N MATHEMATICS

oor

umar

Wasan

rakash

ss and Economics *by R.S. Soni*

Kapoor

n, Benson Town Post,
Tel.: 080-3534673
ennai-600112 Tel.: 044-25322333
16 Tel.: 0141-2282440

Fundamentals of MATHEMATICAL STATISTICS

[Covering the Complete Syllabus of B.A./B.Sc. (Hons. & Pass) Classes of Indian Universities]

By

N.M. Kapoor, Ph.D.

Deptt. of Mathematics
Hans Raj College
University of Delhi

Pitambar Publishing Company Pvt. Ltd.

888, East Park Road, Karol Bagh,
New Delhi-110005 (INDIA)

Published by

PITAMBAR PUBLISHING COMPANY PVT. LTD.

888, East Park Road, Karol Bagh,
New Delhi-110005 (INDIA)
Telephones : 23522997, 23670067, 23625528
Fax : 91-11-23676058
E-mail : pitambar@bol.net.in

Edition
First, 2005

Price
Rs. 250/-

Code No.
29769

ISBN
81-209-1326-4

Printed at

Piyush Printers Publishers Pvt. Ltd.,
New Delhi-110 041.

PREFACI

Statistics today is an indis
forms part of the syllabus of
competitive examinations. Enti

The purpose of this book
rigorously, avoiding non-mathe
of the subject on the part of the
simplicity of exposition as will

Although the textual mate
development of the underlying

The subject is presented in
core of introductory material w
complex and the intricate.

The examples are accompa
of them serve as additional illu
One purpose of the examples
previously treated examples sh
once they are formulated in a na
utility of the book, thought p
arranged are added at the end
various universities are an add

It is hoped that book will
of the subject.

The author would like to
suggestions. Author's thanks a
material and help.

I take this opportunity to
whom I have drawn an inspira
Mathematics for the kind help

The author is also ideble
endeaver.

Suggestions to further i
acknowledged.

EC 2004
Y PVT. LTD.
25528
author
Ltd.,

PREFACE TO THE FIRST EDITION

Statistics today is an indispensable part of every human activity. The subject, therefore, forms part of the syllabus of degree and post-degree as well as several professional and competitive examinations. Entire text is designed to eater to the syllabi of indian universities.

The purpose of this book is to treat statistics as a self-contained mathematical subject rigorously, avoiding non-mathematical concepts. The book assumes no previous knowledge of the subject on the part of the reader and aims at complete clarity for the beginner and such simplicity of exposition as will make the text practically self-teaching.

Although the textual material is concise and to the point, attention has been paid to the development of the underlying concepts. A serious attempt has been made to unify methods.

The subject is presented in a modulated and graded manner beginning with a fundamental core of introductory material which develops gradually from the simple and the easy to the complex and the intricate.

The examples are accompanied by problems. Some of them are simple exercises but most of them serve as additional illustrative material to the text or contain various complements. One purpose of the examples and problems is to develop the reader's intuition. Several previously treated examples show that apparently difficult problems may become almost trite once they are formulated in a natural way and put into proper context. To enhance further the utility of the book, thought provoking questions, carefully selected and systematically arranged are added at the end of each chapter. Problems drawn from latest examinations of various universities are an added attraction of the book.

It is hoped that book will prove to be of much utiliy to the students as well as teachers of the subject.

The author would like to express his appreciation to Mr. V. S. Prasad for his useful suggestions. Author's thanks are also for Miss. Neeru Kapoor for her carefully reading the material and help.

I take this opportunity to thank the various well-known authorities of the world from whom I have drawn an inspiration. I am also grateful to all my colleagues in the Deptt. of Mathematics for the kind help they have given in the preparation of this book.

The author is also idebled to publishers and printers for their co-operation in this endeaver.

Suggestions to further improvement are welcome and will be most gratefully acknowledged.

N.M. KAPOOR

To
My Wife

CONTENTS

Cut Here

Request to t

Please find time

further improvin

improvement wil

FEED-BACK

NAME OF THE BOO

1. Does the bo

2. Is it simple a

3. Are the expl

satisfactory

4. Do you find

useful and s

5. Are the exerc

chapter adeq

6. Are the chart

adequate and

7. Does the boo

and meet you

8. How would y

9. Are there any

book ? ______

Request to the Readers

Please find time to fill the following questionaire to help us in further improving the quality of the book. Suggestions for improvement will be gratefully received and acknowledged.

FEED-BACK QUESTIONAIRE

Name of the Book ____________________

	Yes	No
1. Does the book cover the entire syllabus ?	☐	☐
2. Is it simple and easy to understand ?	☐	☐
3. Are the explanations adequate, clear and satisfactory ?	☐	☐
4. Do you find the chapterwise summaries useful and satisfactory ?	☐	☐
5. Are the exercises given at the end of the chapter adequate ?	☐	☐
6. Are the charts, tables, diagrams, illustrations adequate and satisfactory ?	☐	☐
7. Does the book come upto your expectations and meet your requirements ?	☐	☐

8. How would you rate the book ?

Outstanding	Good
☐	☐
Satisfactory	Poor
☐	☐

9. Are there any factual inaccuracies or tying errors in the book ? ____________________

Cut Here

10. Which sections/chapters you find difficult to understand/ confusing ?

11. Which topics of the syllabus have not been discussed/ inadequately discussed/ discussed in a very elementary or too advanced manner ?

12. What improvements would you like to suggest in the book regarding :

   a) Language and presentation :

   b) Data, explanations, tables, charts, diagrams :

   c) Summaries :

   d) any other suggestion for improvement :

Thanking you for answering the above questionnaire.

P.S. *You can mail your reply to :*

PITAMBAR PUBLISHING CO. PVT. LTD.

888, East Park Road, Karol Bagh,
New Delhi-110 005

Telephones : 23670067, 23522997, 23625528
Fax : 91-11-23676058, 91-11-25765754
E-mail : pitambar@bol.net.in

0.1. The word statistics an
meanings. To a man in the stree
the number of things. To the
physicist statistical is opposite
been derived from Status (Lati
which means an "organized pol
of factual details concerning a
State Craft".

The word 'statistics' whe
orderly manner with specific e
science with techniques which
data.

Statistics is both a scienc
branches of science, are basic
that their successful applicatio
of the statistician.

0.2. Scope of Statistics

The applications of statis
difficult to define its scope but
branch of human knowledge.
rely upon the application of stat
the effects of two drugs, the la
of the eye are exposed to light
electrocardiogram.

Statistics is also used in ag
Various Agricultural problems
for collection, analysis and the

Statistics is an important
of applied mathematics which s
in the defence strategies to pl
day-to-day functioning Gover
used. Calculation of premium
theory of probability and expe

0.3. Distrust of Statistics

Inspite of the very valua
misgiving in the minds of a few

0

Introductory

0.1. The word statistics and the related words 'statistical' and 'statistician' have various meanings. To a man in the street statistics are only figures and statistician as one who counts the number of things. To the economist, statistical stands for the quantitative and to the physicist statistical is opposite of individualistic or exact. The word statistics have either been derived from Status (Latin word) or Statistica (Italian) or Statistik (German) each of which means an "organized political state". Originally statistics was related to the collection of factual details concerning a state and this is why earlier it was known as the "Science of State Craft".

The word 'statistics' when used in plural, means the numerical data collected in an orderly manner with specific end in view but in its singular meaning it means the theoretical science with techniques which deals in to collect, analyse and draw conclusions from the data.

Statistics is both a science and an art. It is a science because its methods, like other branches of science, are basically systematic and have general application and is an art in that their successful application depends to a considerable degree on the skill and experience of the statistician.

0.2. Scope of Statistics

The applications of statistics are so numerous and ever-increasing that not only it is difficult to define its scope but also unwise to do so. These days it is introducing into every branch of human knowledge. A student of science, while conducting an experiment, has to rely upon the application of statistics. In biology, testing of significance is applied to compare the effects of two drugs, the law of probability is used in radiation when the cells in the retina of the eye are exposed to light and statistical papers are used to study heart-beats through electrocardiogram.

Statistics is also used in agriculture and used in this way it is called *Agricultural Statistics*. Various Agricultural problems are solved or simplified by applying some suitable scheme for collection, analysis and the interpretation.

Statistics is an important member of the mathematical family. It is regarded as a branch of applied mathematics which specialises in data. Statistical techniques are of great assistance in the defence strategies to plan maximum destruction with minimum effect. Also for the day-to-day functioning Government depends upon statistics. In Insurance also statistics is used. Calculation of premiums and annuity etc., is wholly a statistical work based on the theory of probability and expectation.

0.3. Distrust of Statistics

Inspite of the very valuable service that statistics renders, there is some amount of misgiving in the minds of a few people with regard to its reliability and usefulness. It is said :

(1) "With statistics anything can be proved".

(2) There are three kinds of lies : namely (*i*) lies (*ii*) damned lies and (*iii*) statistics-wicked in the order of their meaning.

0.4. Limitations of Statistics

Statistics has its own limitations. It cannot be applied to all kinds of phenomena and cannot be made to answer all queries. Few of its main limitations are listed below :

(1) It deals only with those subjects of inquiry which can be measured quantitatively and can be expressed numerically.

(2) It deals only with aggregates of facts and no importance is attached to individual items.

(3) Statistical data is only approximately and not mathematically correct.

(4) Statistics can be used to establish wrong conclusions and therefore can be used only by experts.

The methods by which statistical data are analyzed are called *statistical methods.* These methods range from the most elementary descriptive devices which may be understood by the common man to those complicated mathematical procedures which can be apprehended only by the expert theoreticians. The mathematical theory which is the basis of these methods is called the *theory of statistics* or *mathematical statistics.* The purpose of this text is to discuss the fundamental principles and theory of statistics in simple and easily comprehensible manner.

□□

Frequency D
C

1.1. Introduction

By a variable is meant a
continuous or discrete. If a va
called a continuous variable. .
values, is called discrete. Hei
test scores, *etc.*, are discrete v

In this and next few Cha
results obtained will also be t

1.2. Frequency Distribution

A frequency distributio
times each value is taken a
called the *frequency* of tha
frequency.

When the values of t
representation is called **grou**
and the boundary figures are
limits and those on the righ
limits is called class-interva
class frequency. The value
values or **central values.**

For computational purpo
frequency distribution can be

where $x_1, x_2 \ldots, x_n$ are the va

Let $f_1 + f_2 + \ldots + f_n = N$

Then, N is total frequenc

1.3. Graphic Representation

It is one of the most convi
a bird's eye-view of entire da
These help one in making qui

ned lies and (*iii*) statistics-

ll kinds of phenomena and
ıs are listed below :
ɔe measured quantitatively

:e is attached to individual

ically correct.
. therefore can be used only

l *statistical methods*. These
ıich may be understood by
which can be apprehended
ȝ the basis of these methods
: purpose of this text is to
: and easily comprehensible

❑❑

1

Frequency Distribution and Measures of Central Tendency

1.1. Introduction

By a variable is meant a quantity which assumes different values. A variable may be continuous or discrete. If a variable can take any numerical value within a certain range it is called a continuous variable. A variable, for which there is a gap between its two successive values, is called discrete. Heights, weights are examples of continuous variable and marks, test scores, *etc.*, are discrete variables.

In this and next few Chapters only discrete variables will be considered although the results obtained will also be true for continuous variables.

1.2. Frequency Distribution

A frequency distribution is one where the values of the variable and the number of times each value is taken are put together. The number of times the value is taken is called the *frequency* of that value. The sum of all the frequencies is called **total frequency.**

When the values of the variable are presented in the form of groups, the representation is called **grouped frequency distribution.** The groups are called **classes** and the boundary figures are called **class limits.** The figures on the left are called **lower limits** and those on the right are called **upper limits.** The difference between the two limits is called class-interval or width of the class. The frequency of the class is called **class frequency.** The values midway between lower and upper limits are called **mid-values** or **central values.**

For computational purposes, it is assumed that variate takes mid-values only. Thus, a frequency distribution can be taken in the form

$$\begin{array}{llll} x \rightarrow x_1 & x_2 & \ldots\ldots\ldots\ldots & x_n \\ f \rightarrow f_1 & f_2 & \ldots\ldots\ldots\ldots & f_n \end{array}$$

where $x_1, x_2 \ldots, x_n$ are the values of the variable x with frequencies $f_1, f_2, \ldots, f_n$.

Let $f_1 + f_2 + \ldots + f_n = N$

Then, N is total frequency.

1.3. Graphic Representation of a Frequency Distribution

It is one of the most convincing and appealing method of presenting a data. Graphs give a bird's eye-view of entire data and therefore information presented is easily understood. These help one in making quick and accurate comparison of data.

A frequency distribution can be presented graphically in any of the following ways :

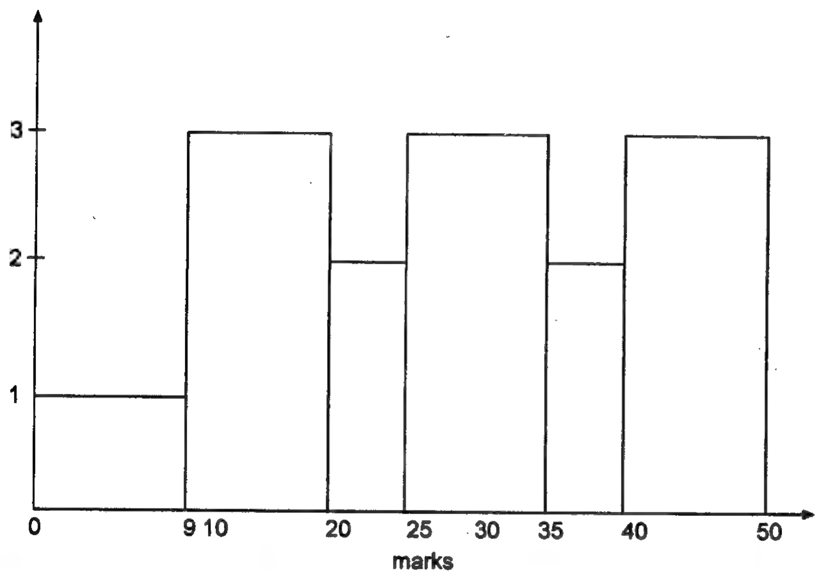
(*i*) Histogram

It is a set of adjacent rectangles whose areas are proportional to the frequencies.

To draw histogram, class-intervals are marked along x-axis on a suitable scale. On the class-intervals as bases rectangles are drawn with heights proportional to the ratio of corresponding frequency and class-width. For equal class-intervals the heights of the rectangles will be proportional to the frequencies.

The figure below exhibits the histogram of the following frequency distribution :

Marks	*No. of students*	*Marks*	*No. of students*
0–9	9	25–35	30
9–20	33	35–40	10
20–25	10	40–50	30



The x-axis, of course, need not start with zero but the y-axis must start with zero and must have no scale break. Sometimes a space is left between the first rectangle and the vertical axis. The base of each rectangle is labelled on both sides in terms of the class-limits if the data is continuous (*i.e.*, the upper limit of one class and the lower limit of the following class coincide). In discontinuous data, only the lower limits are marked on x-axis. Some statisticians, however, in presenting discontinuous data, leave small gaps between the rectangles and label both limits of each class. Another way of labelling the horizontal axis on the histogram is by showing the mid-value in the middle of the base of the rectangle.

The histogram sometimes is also used even for representing ungrouped data. Here different values are regarded as the mid-points of the classes and corresponding frequencies are supposed to spread over the whole class.

However, for data with open-end classes histogram cannot be constructed. One solution to this problem is to plot the histogram without the open-end class or classes and to add the information concerning them in figures.

(*ii*) Frequency Polygon

It is used to represent the grouped or ungrouped frequency distribution. For a grouped frequency distribution mid-points of the classes are taken as variate values and frequency of

Define Frequ

the class is taken as frequenc

Points $(x_1, f_1), (x_2, f_2)$,
are joined by straight lines.
The point (x_1, f_1) is joi
dotted lines.
The polygon so obtaine
For equal class-intervals
points of the upper sides of
lines.
Frequency polygons for
graph which is not possible f
graphic comparison of vario
Thus, for reading the
histogram presents a far be
frequency polygon.

(*iii*) Frequency Curve

If the number of observ
reduced without noting dow
frequency polygon will go on
curve can be obtained of pas
The larger the number
consequently the correspond
to become a smooth curve.
This smooth curve is c
polygon or various tops of th
area under the curve is appro
as regular as possible and al
end at the base line and as a
just outside the histogram.
The area under the curv

(*iv*) Cumulative Frequenc

Here cumulative frequ
frequencies are obtained by
top. Points are then plotted
corresponding cumulative f
straight lines.
The graph so obtained is
Curve is obtained by appro
The cumulative frequen
The cumulative freque
(more) than type. These ar
obtained is called **less (mo**
(finishing) point of this cur
class as abscissa).

Define Frequency Polygon.

the class is taken as frequency of the mid-point. Thus, the data is obtained in the form :

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ f_1 & f_2 & \cdots & f_n \end{pmatrix}$$

Points $(x_1, f_1), (x_2, f_2), \ldots (x_n, f_n)$ are plotted on the graph paper and successive points are joined by straight lines.

The point (x_1, f_1) is joined with $(x_0, 0)$ and the point (x_n, f_n) with $(x_{n+1}, 0)$ through dotted lines.

The polygon so obtained is called **frequency polygon**.

For equal class-intervals, the frequency polygon can be obtained by joining the middle points of the upper sides of the adjacent rectangles of the histogram by means of straight lines.

Frequency polygons for two or more frequency distributions can be shown on a single graph which is not possible for histograms. Hence, frequency polygons are preferred for the graphic comparison of various frequency distributions.

Thus, for reading the relationship of individual class frequencies to the total, histogram presents a far better picture and as such is preferred than the corresponding frequency polygon.

(iii) Frequency Curve

If the number of observations goes on increasing, the size of the class-interval can be reduced without noting down a fall in the individual class frequencies. The points of the frequency polygon will go on becoming nearer horizontally and hence an approximate smooth curve can be obtained of passing through most of these points.

The larger the number of observations would make the class intervals smaller and consequently the corresponding frequency polygon would approach more and more closely to become a smooth curve.

This smooth curve is called **frequency curve.** It is obtained by smoothing either the polygon or various tops of the histogram. The curve is drawn freehand in such a manner that area under the curve is approximately the same as that of the polygon. The curve should look as regular as possible and all sudden turns should be avoided. The curve should begin and end at the base line and as a general rule it may be extended to the mid-points of the classes just outside the histogram.

The area under the curve should represent the total frequency.

(iv) Cumulative Frequency Curve or Ogive

Here cumulative frequencies are used and not the class frequencies. Cumulative frequencies are obtained by adding frequencies either from top to bottom or from bottom to top. Points are then plotted with upper (or lower) limits of the classes as abscissae and corresponding cumulative frequency as ordinates and successive points are then joined by straight lines.

The graph so obtained is called **Cumulative frequency polygon, Cumulative frequency Curve** is obtained by approximating the polygon through the smooth curve.

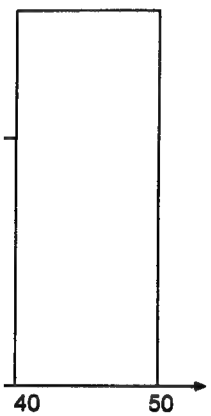
The cumulative frequency curve is also called **ogive.**

The cumulative frequencies taken from top (bottom) to bottom (top) are called **less (more) than type.** These are plotted against upper (lower) class limits. Polygon or curve so obtained is called **less (more) than type** cumulative frequency polygon or curve. Starting (finishing) point of this curve would be the origin (point on x-axis with upper limit of last class as abscissa).

ıy of the following ways :

ıal to the frequencies.
on a suitable scale. On the
roportional to the ratio of
ıtervals the heights of the

equency distribution :
No. of students
30
10
30



·axis must start with zero
een the first rectangle and
oth sides in terms of the
one class and the lower
only the lower limits are
discontinuous data, leave
ıch class. Another way of
e mid-value in the middle

ing ungrouped data. Here
corresponding frequencies

constructed. One solution
ss or classes and to add the

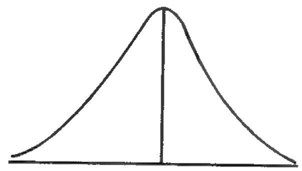
listribution. For a grouped
ıte values and frequency of

These curves are frequently used in estimating various partition values *e.g.*, if a line corresponding to $\frac{N}{4}$ is drawn parallel to x-axis, first quartile will be obtained. Similarly, the other partition values can be obtained.

If instead of cumulative frequencies percentage cumulative frequencies (*i.e.*, cumulative frequency expressed as percentage of the total frequency) are taken, curve obtained is called **percentage** cumulative frequency curve. This is useful in comparing different frequency distributions.

Remark. Some special types of curves which generally we come across in statistics are :

(*i*) Symmetrical curves such curves can be folded along a vertical line so that two halves of the curve coincide. Such curves arise for a distribution in which class-frequencies go on decreasing symmetrically on either side of central value.



Symmetrical Curve

(*ii*) Moderately asymmetrical or skewed curves

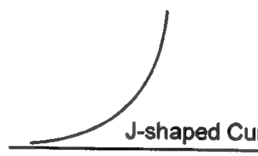
Curves which are not symmetrical are called skewed. Such curves arise for a distribution in which frequencies decrease with significant rapidity on one side of the maximum than on the other.

(*iii*) Extremely asymmetrical (or J-shaped) curves

Such curves are of two shapes :

J-shape and τ-shape (reverse J-shape).

These curves occur for distribution in which frequency is maximum at one end of value and decrease to minimum to other end.



J-shaped Curve

τ-shaped Curve

(*iv*) U-shaped curves

These curves are of shape U and occur for a distribution in which frequencies increase as we move from centre towards ends of values.

U-shaped Curve

1.4. Measures of location or central tendency

These are statistical constants which give an idea about the concentration of the values in the central part of the distribution. It can be thought of as the value of the variable which is representative of the entire distribution. The following are the various measures of central tendency.

(*i*) Arithmetic Mean. *It is defined by*

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots\dots + x_n f_n}{f_1 + f_2 + \dots\dots + f_n}$$

To obtain it for a given data, calculations are simplified by taking the variate defined by

$$u = \frac{x-a}{h}$$

where 'a' and 'h' are to be chose

where $\bar{u}$ is A.M. of u.

***Ex. 1-1.** Show that the algel arithmetic mean is zero.*

Sol. Let $x_1, x_2 \dots, x_n$ be the

Let $N = f_1 +$

Let $\bar{x}$ be A.M

Now $\sum_{i=1}^{n} f_i (x_i$

Ex. 1-2. *Show that the arithi*

Sol.

Ex. 1-3. *From the data give*

S.N.
1
2
3
4
5
6
7
8
9

Sol. (*i*) *Direct Method.*

Arithmetic Mean (A.M.) =

$= \frac{17+32+35+33+15+21}{}$

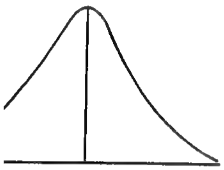
$= \frac{431}{18} = 23 \cdot 94.$

partition values *e.g.*, if a line

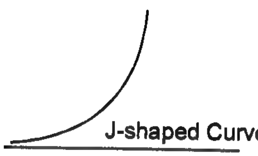
will be obtained. Similarly, the

ve frequencies (*i.e.*, cumulative
taken, curve obtained is called
comparing different frequency

y we come across in statistics

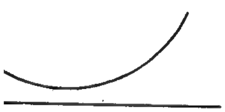


mmetrical Curve



J-shaped Curve

τ-shaped Curve



l-shaped Curve

concentration of the values
value of the variable which
various measures of central

fied by taking the variate

where 'a' and 'h' are to be chosen suitably. Then A.M. is given by

$$\bar{x} = a + h\bar{u}$$

where $\bar{u}$ *is A.M. of u.*

Ex. 1-1. *Show that the algebraic sum of the deviations of a set of values from their arithmetic mean is zero.*

Sol. Let $x_1, x_2 \ldots, x_n$ be the set of values with frequencies $f_1, f_2 \ldots f_n$.

Let $\quad N = f_1 + f_2 + \ldots\ldots + f_n$

Let $\quad \bar{x}$ be A.M. Then $\bar{x} = \dfrac{1}{N}\sum_{i=1}^{n} f_i x_i$

Now
$$\sum_{i=1}^{n} f_i(x_i - x) = \sum_{i=1}^{n} f_i x_i - \bar{x}\sum_{i=1}^{n} f_i$$
$$= N\bar{x} - \bar{x}N$$
$$= 0$$

Ex. 1-2. *Show that the arithmetic mean of the first n natural numbers is* $\dfrac{n+1}{2}$.

Sol.
$$\text{A.M.} = \frac{1+2+\ldots+n}{n}$$
$$= \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

Ex. 1-3. *From the data given below, calculate mean :*

S.N.	*Marks*	*S.N.*	*Marks*
1	17	10	18
2	32	11	20
3	35	12	22
4	33	13	11
5	15	14	15
6	21	15	35
7	41	16	23
8	32	17	38
9	11	18	12

Sol. (*i*) *Direct Method.*

$$\text{Arithmetic Mean (A.M.)} = \frac{\Sigma x}{n}$$

$$= \frac{17+32+35+33+15+21+41+32+11+18+20+22+11+15+35+23+38+12}{18}$$

$$= \frac{431}{18} = 23\cdot 94.$$

(ii) Short-cut method

S.N.	x Marks	$x-23=u$	S.N.	x Marks	$x-23=u$
1	17	−6	10	18	−5
2	32	9	11	20	−3
3	35	12	12	22	−1
4	33	10	13	11	−12
5	15	−8	14	15	−8
6	21	−2	15	35	12
7	41	18	16	23	0
8	32	9	17	38	15
9	11	−12	18	12	−11
					17

$$\therefore \quad \text{A.M.} = 23 + \frac{\Sigma u}{n} = 23 + \frac{17}{18}$$

$$= \frac{414+17}{18} = \frac{431}{18} = 23 \cdot 94.$$

Ex. 1-4. *Calculate the mean from the following data :*

Size	Frequency	Size	Frequency
4—8	6	24—28	12
8—12	10	28—32	10
12—16	18	32—36	6
16—20	30	36—40	2
20—24	15		

Sol.

Size	Frequency f	Mid points x	$u = \frac{x-22}{4}$	uf
4—8	6	6	−4	−24
8—12	10	10	−3	−30
12—16	18	14	−2	−36
16—20	30	18	−1	−30
20—24	15	22	0	0
24—28	12	26	1	12
28—32	10	30	2	20
32—36	6	34	3	18
36—40	2	38	4	8
	109			−62

$$\therefore \quad \text{A.M.} = 22 + \left(\frac{\Sigma fu}{N}\right) \times 4 = 22 - \frac{62}{109} \times 4$$

$$= 19 \cdot 725.$$

Ex. 1-5. *Calculate the mea*

Income between (in Rs.)	No.
100—200	
100—300	
100—400	

The data is given in the f
mean, it is to be converted into
Now the number of person
persons having income betwee
∴ The number of persons
In a similar manner the f
calculated.

Income	Given Freq. (c.f.)
100—200	15
200—300	33
300—400	63
400—500	83
500—600	100

$$\therefore \text{A.M.} = 350 + \frac{6}{100} \cdot 100$$

$$= 356.$$

Ex. 1-6. *Calculate the mea*

Marks	*No. o*
More than 0	
More than 10	
More than 20	
More than 30	

Sol. The data is given in cu
first converted into ordinary fre
Now the number of student

and the number of students getti

∴ The number of students

Proceeding likewise the fr
obtained.
In the end since there is no s
will be 60—70.
Thus, we have the table :

x Marks	x–23 = u
18	–5
20	–3
22	–1
11	–12
15	–8
35	12
23	0
38	15
12	–11
	17

·94.

Frequency
12
10
6
2

$= \frac{x-22}{4}$	uf
–4	–24
–3	–30
–2	–36
–1	–30
0	0
1	12
2	20
3	18
4	8
	–62

$\frac{}{9} \times 4$

Ex. 1-5. *Calculate the mean from the following data :*

Income between (in Rs.)	*No. of persons*	*Income between (in Rs.)*	*No. of persons*
100—200	15	100—500	83
100—300	33	100—600	100
100—400	63		

The data is given in the form of cumulative frequency distribution. For calculating mean, it is to be converted into an ordinary frequency distribution.

Now the number of persons having income between 100—200 = 15 and the number of persons having income between 100—300 = 33.

∴ The number of persons having income between 200—300 is 33 – 15 = 18.

In a similar manner the frequencies of groups 300—400, 400—500 *etc.*, can be calculated.

Income	*Given Freq. (c.f.)*	*Frequency (f)*	*(x) Mid pt.*	$u = \left(\frac{x-350}{100}\right)$	*uf*
100—200	15	15	150	–2	–30
200—300	33	33–15 = 18	250	–1	–18
300—400	63	63–33 = 30	350	0	0
400—500	83	83–63 = 20	450	1	20
500—600	100	100–83 = 17	550	2	34
		100			6

$$\therefore \text{ A.M.} = 350 + \frac{6}{100} \cdot 100$$
$$= 356.$$

Ex. 1-6. *Calculate the mean for the data given below :*

Marks	*No. of students*	*Marks*	*No. of students*
More than 0	100	More than 40	25
More than 10	90	More than 50	15
More than 20	75	More than 60	5
More than 30	50	More than 70	0

Sol. The data is given in cumulative distribution type. For calculating mean it is to be first converted into ordinary frequency distribution.

Now the number of students getting marks more than

'0' = 100

and the number of students getting marks more than

'10' = 90

∴ The number of students getting marks between

0—10 = 100 – 90 = 10.

Proceeding likewise the frequencies of other classes 10—20, 20—30 *etc.*, can be obtained.

In the end since there is no student getting marks more than 70, the last class in the table will be 60—70.

Thus, we have the table :

Note. Classes below are closed from right.

Sol.

Marks	*Given Freq.*	*Frequencies (f)*	*Mid-pt. (x)*	$u = \dfrac{x-35}{10}$	*uf*
0–10	100	$100 - 90 = 10$	5	–3	–30
10–20	90	$90 - 75 = 15$	15	–2	–30
20–30	75	$75 - 50 = 25$	25	–1	–25
30–40	50	$50 - 25 = 25$	35	0	0
40–50	25	$25 - 15 = 10$	45	1	10
50–60	15	$15 - 5 = 10$	55	2	20
60–70	5	5	65	3	15
		100			–40

$$\therefore \qquad \text{A.M.} = 35 + \frac{(-40)}{100} \times 10$$
$$= 31.$$

Ex. 1-7. *The following table gives the frequency distribution of monthly salaries of 70 employees of company X.*

Salary (in Rs.)	*No. of Employees*
100—119	8
120—139	10
140—159	16
160—179	15
180—199	10
200—239	8
240—259	3
	70

Compute the arithmetic mean.

Sol. The class intervals are given in inclusive forms. A value less than 119·5 will be counted in first class and a value greater than or equal to it in the second, so for calculation the class 100—119 is as good as 99·5—119·5 and so on. Thus, we have the following table :

Salary	*Frequency (f)*	*Mid value (x)*	$d = \left(\dfrac{x-169{\cdot}5}{10}\right)$	*df*
99·5—119·5	8	109·5	–6	–48
119·5—139·5	10	129·5	–4	–40
139·5—159·5	16	149·5	–2	–32
159·5—179·5	15	169·5	0	0
179·5—199·5	10	189·5	2	20
199·5—239.5	8	219·5	5	40
239·5—259·5	3	249·5	8	24
	70			–36

$$\therefore \qquad \text{A.M.} = 169{\cdot}5 - \frac{36}{70} \times 10$$

Ex. 1-8. *Illustrate by an exa*
(i) adding 'a' to every
(ii) subtracting 'a' fron
(iii) Multiplying every i
(iv) Dividing every item
on the arithmetic mean of a

Sol. Let the series be

$\therefore$ A

(*i*) By adding '*a*' to every it
2 +

$\therefore$ A

so that arithmetic mean also inc
(*ii*) By subtracting '*a*' from
2

$\therefore$ A

so that arithmetic mean also dim
(*iii*) By multiplying each ite

$\therefore$ A

so that arithmetic mean is also n
(*iv*) By dividing each item b

$\therefore$ A

so that arithmetic mean is also d

Ex. 1-9. *If m_1, m_2 be the ari*
Find the A.M. of the series obta

)	$u = \dfrac{x-35}{10}$	uf
	−3	−30
	−2	−30
	−1	−25
	0	0
	1	10
	2	20
	3	15
		−40

ıtion of monthly salaries of 70

Employees
8
10
16
15
10
8
3
70

. value less than 119·5 will be
ı the second, so for calculation
s, we have the following table :

$d = \left(\dfrac{x-169\cdot5}{10}\right)$	df
−6	−48
−4	−40
−2	−32
0	0
2	20
5	40
8	24
	−36

$$= 169\cdot5 - \frac{36}{7}$$

$$= 169\cdot5 - 5\cdot143 = 164\cdot357.$$

Ex. 1-8. *Illustrate by an example the effect of*

(i) adding 'a' to every item.
(ii) subtracting 'a' from every item.
(iii) Multiplying every item by 'a'.
(iv) Dividing every item by 'a'

on the arithmetic mean of a series.

Sol. Let the series be

$$2, 3, 4, 5, 6$$

$$\therefore \qquad \text{A.M.} = \frac{2+3+4+5+6}{5} = \frac{20}{5} = 4$$

(*i*) By adding 'a' to every item, the new series is

$$2+a,\ 3+a,\ 4+a,\ 5+a,\ 6+a$$

$$\therefore \qquad \text{A.M.} = \frac{(2+a)+(3+a)+(4+a)+(5+a)+(6+a)}{5}$$

$$= 4+a$$

so that arithmetic mean also increases by 'a'.

(*ii*) By subtracting 'a' from each item, the new series is

$$2-a, 3-a, 4-a, 5-a, 6-a$$

$$\therefore \qquad \text{A.M.} = \frac{(2-a)+(3-a)+(4-a)+(5-a)+(6-a)}{5}$$

$$= 4-a$$

so that arithmetic mean also diminishes by 'a'.

(*iii*) By multiplying each item by 'a', the new series is

$$2a, 3a, 4a, 5a, 6a$$

$$\therefore \qquad \text{A.M.} = \frac{2a+3a+4a+5a+6a}{5}$$

$$= 4a$$

so that arithmetic mean is also multiplied by 'a'.

(*iv*) By dividing each item by 'a', the new series is

$$\frac{2}{a}, \frac{3}{a}, \frac{4}{a}, \frac{5}{a}, \frac{6}{a}.$$

$$\therefore \qquad \text{A.M.} = \frac{1}{5}\left(\frac{2}{a}+\frac{3}{a}+\frac{4}{a}+\frac{5}{a}+\frac{6}{a}\right)$$

$$= \frac{4}{a}$$

so that arithmetic mean is also divided by 'a'.

Ex. 1-9. *If m_1, m_2 be the arithmetic means for two series of sizes n_1 and n_2 respectively. Find the A.M. of the series obtained on combining them.*

Sol. Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2 \cdots y_{n_2}$ be the items of two series respectively. Then

$$m_1 = \frac{1}{n_1}\left(x_1 + x_2 + \ldots + x_{n_1}\right)$$

$$m_2 = \frac{1}{n_2}\left(y_1 + y_2 + \ldots + y_{n_2}\right)$$

Let m be the A.M. of the combined series.

Then
$$m = \frac{1}{n_1 + n_2}\left\{\left(x_1 + x_2 + \ldots + x_{n_1}\right) + \left(y_1 + y_2 + \ldots + y_{n_2}\right)\right\}$$

$$= \frac{1}{n_1 + n_2}\left\{m_1 n_1 + m_2 n_2\right\}.$$

Ex. 1-10. *The mean of the marks obtained in an examination by a group of 100 students was found to be 49·46. The mean of the marks obtained in the same examination by another group of 200 students was 52·32. Find the mean of the marks obtained by both the groups of students taken together.*

Sol. Here $m_1 = 49 \cdot 46, \quad n_1 = 100$

$m_2 = 52 \cdot 32, \quad n_2 = 200$

$$\therefore \quad m = \frac{4946 + 10464}{100 + 200}$$

$$= \frac{15410}{300}$$

$$= 51 \cdot 37.$$

Ex. 1-11. *Two groups of students reported mean weights of 162 and 148 pounds respectively. When would the mean weights of both groups together be 155 pounds?*

Sol. Here $m = 155, m_1 = 162, m_2 = 148$. Let n_1 and n_2 be the sizes of two groups.

Then,
$$m = \frac{n_1 m_1 + m_2 n_2}{n_1 + n_2}$$

$$155 = \left(\frac{n_1}{n_1 + n_2}\right) 162 + \left(\frac{n_2}{n_1 + n_2}\right)(148)$$

$$= \left(\frac{n_1}{n_1 + n_2}\right) 162 + \left(1 - \frac{n_1}{n_1 + n_2}\right)(148)$$

$$= \left(\frac{n_1}{n_1 + n_2}\right)(162 - 148) + 148$$

$$\therefore \quad 14\left(\frac{n_1}{n_1 + n_2}\right) = 155 - 148 = 7$$

$$\therefore \quad \frac{n_1}{n_1 + n_2} = \frac{7}{14} = \frac{1}{2}$$

$$\therefore \qquad 2n_1 = n_1 + n_2$$

$$\text{or} \qquad n_1 = n_2$$

$\therefore$ Two groups must be of same size.

Ex. 1-12. *The mean annual salary paid to all employees of a company was Rs. 5000. The mean annual salaries paid to male and female employees were Rs. 5200 and Rs. 4200 respectively. Determine the percentage of males and females employed by the company.*

Sol. Here $m = 5000,\ m_1 = 5200, m_2 = 4200$. Let n_1 and n_2 be the number of male and female employees.

Now,
$$m = \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2}$$

$$\therefore \qquad 5000 = \left(\frac{n_1}{n_1 + n_2}\right)(5200) + \left(\frac{n_2}{n_1 + n_2}\right)(4200)$$

$$\text{or} \qquad 50 = \left(\frac{n_1}{n_1 + n_2}\right)(52) + \left(1 - \frac{n_1}{n_1 + n_2}\right)(42)$$

$$= \left(\frac{n_1}{n_1 + n_2}\right)(52 - 42) + 42$$

$$\text{or} \quad \left(\frac{n_1}{n_1 + n_2}\right)(10) = 8$$

$$\text{or} \qquad \frac{n_1}{n_1 + n_2} = \frac{8}{10}$$

$$\therefore \qquad \frac{n_2}{n_1 + n_2} = 1 - \frac{n_1}{n_1 + n_2} = 1 - \frac{8}{10} = \frac{2}{10}$$

$\therefore$ no. of male employees $= 80\%$

and no. of female employees $= 20\%$.

Ex. 1-13. *The population of five towns A, B, C, D, E was 20,000; 26,000; 23,000; 25,000 and 24,000 respectively. The average income of the resident for the respective towns was 280, 270, 240, 230 and 300.*

Find out the average income per head for all the towns combined.

Sol. Let n_1, n_2, n_3, n_4, n_5 be the population of five towns and m_1, m_2, m_3, m_4, m_5 be the average income of the resident for the respective towns. Then

$$n_1 = 20{,}000, n_2 = 26{,}000, n_3 = 23{,}000, n_4 = 25{,}000, n_5 = 24{,}000$$

and
$$m_1 = 280, m_2 = 270, m_3 = 240, m_4 = 230, m_5 = 300$$

Let m be the average income per head for all the towns combined.

Then
$$m = \frac{m_1 n_1 + m_2 n_2 + m_3 n_3 + m_4 n_4 + m_5 n_5}{n_1 + n_2 + n_3 + n_4 + n_5}$$

$$= \frac{5600 + 7020 + 5520 + 5750 + 7200}{20 + 26 + 23 + 25 + 24}$$

$$= \frac{31090}{118} = 263 \cdot 47.$$

Ex. 1-14. *What is the average of daily wages for the workers of the two factories combined :*

	Factory A	*Factory B*
No. of wage earners	250	200
Average daily wage	Rs. 2·00	Rs. 2·50

Sol. Let m be the average of daily wages for the workers of the two factories combined.

Then,
$$m = \frac{(250)(2) + (200)(2 \cdot 5)}{250 + 200}$$

$$= \frac{50 + 50}{45} = \frac{100}{45}$$

$$= \text{Rs. } 2{\cdot}22.$$

Ex. 1-15. *The mean marks of 100 students were found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the corrected mean corresponding to the corrected score.*

Sol. Let x be the variable for the marks.

Then, $\bar{x} = 40$

and $n(=\text{size}) = 100$

$\therefore$ $\frac{1}{n} \Sigma x = 40$

$\therefore$ $\Sigma x = 40 \times 100 = 4000$

Corrected value of $\Sigma x = 4000 - 83 + 53$

$= 3970$

$\therefore$ Corrected mean $= \frac{3970}{100} = 39 \cdot 7$

(*ii*) Geometric and Harmonic Means

Geometric Mean: *It is defined by*

$$G = \left\{ x_1^{f_1} \cdot x_2^{f_2} \ldots . x_n^{f_n} \right\}^{1/N}$$

were $N = f_1 + f_2 + \ldots + f_n$.

For a given data it is obtained by finding A.M. of log x and then taking 'antilog'.

Harmonic Mean : *It is defined by*

$$\frac{1}{H} = \frac{f_1 \cdot \frac{1}{x_1} + f_2 \cdot \frac{1}{x_2} + \ldots . + f_n \cdot \frac{1}{x_n}}{f_1 + f_2 + \ldots . + f_n}$$

where H is the harmonic mean.

For a given data it is obtained by finding A.M. of $\frac{1}{x}$ *and then taking its reciprocal.*

Ex. 1-16. *Find the geome*

Sol. Let x be the variable.

$\therefore$ Values of $y = \log_{10} x$

$0, \log_{10} 2, 2$

$\therefore$ $\log_{10}$ (

$\therefore$

Ex. 1-17. *Calculate Geon*

6·5, 169·0

Sol.

x
6·5
169·0
11·0
112·5
14·2
75·5
35·5
215·0

$\log_{10}$ (G.M.) $= \frac{1}{n}$ (

$\therefore$

orkers of the two factories

Factory B
200
Rs. 2·50
·kers of the two factories

). Later on it was discovered
·responding to the corrected

l then taking 'antilog'.

then taking its reciprocal.

Ex. 1-16. *Find the geometric mean of the series 1, 2, 4, 8, ...2^n.*

Sol. Let x be the variable. Then values of x are

$$1, 2, 4, 8, \ldots\ldots\ldots\ldots 2^n$$

$\therefore$ Values of $y = \log_{10} x$ are

$$0, \log_{10} 2, 2\log_{10} 2, 3\log_{10} 2, \ldots n\log_{10} 2.$$

$$\therefore \qquad \log_{10}(\text{G.M.}) = \frac{0 + \log_{10} 2 + 2\log_{10} 2 + 3\log_{10} 2 + \ldots + n\log_{10} 2}{(n+1)}$$

$$= (\log_{10} 2)\frac{1}{n+1}\{1 + 2 + \ldots + n\}$$

$$= (\log_{10} 2)\frac{1}{n+1}\cdot\frac{n(n+1)}{2}$$

$$= \frac{n}{2}\log_{10} 2$$

$$= \log_{10} 2^{n/2}$$

$$\therefore \qquad \text{G.M.} = 2^{n/2}.$$

Ex. 1-17. *Calculate Geometric and Harmonic means from the following data :*

6·5, 169·0, 11·0, 112·5, 14·2, 75·5, 35·5, 215·0.

Sol.

x	$\log_{10} x$	$\frac{1}{x}$
6·5	0·8129	0·1539
169·0	2·2279	0·0059
11·0	1·0414	0·0909
112·5	2·0512	0·0089
14·2	1·1523	0·0704
75·5	1·8779	0·0133
35·5	1·5502	0·0282
215·0	2·3324	0·0047
	13·0462	0·3762

$$\log_{10}(\text{G.M.}) = \frac{1}{n}(\Sigma \log_{10} x) = \frac{13\cdot0462}{8} = 1\cdot630775$$

$$\cong 1\cdot6308$$

$$\therefore \qquad \text{G.M.} = 42\cdot74$$

$$\frac{1}{\text{H.M.}} = \frac{\left(\sum \frac{1}{x}\right)}{n} = \frac{0\cdot3762}{8}$$

$$\text{H.M.} = 21\cdot27.$$

Ex. 1-18. *Find the geometric mean and harmonic mean for the data of Ex. 1-4.*

Class Intervals	Freq. (f)	Mid points x	$\log_{10} x$	$\frac{1}{x}$	$f.(\log_{10} x)$	$f.\frac{1}{x}$
4—8	6	6	0·778151	0·166667	4·668906	1·000002
8—12	10	10	1·000000	0·100000	10·000000	1·000000
12—16	18	14	1·146128	0·071429	20·630304	1·285722
16—20	30	18	1·255273	0·055556	37·65819	1·666680
20—24	15	22	1·342423	0·045455	20·136345	0·681825
24—28	12	26	1·414973	0·038462	16·979676	0·461544
28—32	10	30	1·477121	0·033333	14·77121	0·333330
32—36	6	34	1·531479	0·029412	9·188874	0·176472
36—40	2	38	1·579784	0·026316	3·159568	0·052632
	109				137·193073	6·658207

$$\therefore \qquad \log_{10} (\text{G.M.}) = \frac{\Sigma f (\log_{10} x)}{N}$$

$$= \frac{137 \cdot 193073}{109} = 1 \cdot 25865 = 1 \cdot 2587$$

$$\therefore \qquad \text{G.M.} = 18 \cdot 14$$

and
$$\frac{1}{\text{H.M.}} = \frac{\left(\sum \frac{f}{x}\right)}{N} = \frac{6 \cdot 658207}{109}$$

$$\therefore \qquad \text{H.M.} = \frac{109}{6 \cdot 658207} = 16 \cdot 37.$$

Ex. 1-19. *In previous Ex. find quadratic mean.*

Class-Intervals	*Frequency (f)*	*Mid-points x*	x^2	fx^2
4—8	6	6	36	216
8—12	10	10	100	1000
12—16	18	14	196	3528
16—20	30	18	324	9720
20—24	15	22	484	7260
24—28	12	26	676	8112
28—32	10	30	900	9000
32—36	6	34	1156	6936
36—40	2	38	1444	2888
	109			48660

$$\therefore \qquad \text{Quadratic mean} = \frac{\left(\sum fx^2\right)}{N}$$

$$= \frac{48660}{109} = 446 \cdot 42.$$

Ex. 1-20. *If g_1 and g_2 b*
the geometric mean of the se

Sol. Let $x_1, x_2 ... x_{n_1}$ and

Then,

and

Let g be the G.M. of the

Then,

Ex. 1-21. *If $x_1, x_2 ... x_n$*
H.M. 'H', show that

Sol.

and

From inequalities

∴

Also $\frac{1}{x_1}$

∴ $\frac{1}{H} \geq$

∴ $A \geq$

· the data of Ex. 1-4.

. ($\log_{10} x$)	$f.\frac{1}{x}$
4·668906	1·000002
0·000000	1·000000
0·630304	1·285722
37·65819	1·666680
0·136345	0·681825
6·979676	0·461544
4·77121	0·333330
9·188874	0·176472
3·159568	0·052632
37·193073	6·658207

$= 1 \cdot 2587$

x^2	fx^2
36	216
100	1000
196	3528
324	9720
484	7260
676	8112
900	9000
1156	6936
1444	2888
	48660

Ex. 1-20. *If g_1 and g_2 be the geometric means of two series of n_1 and n_2 items. Find the geometric mean of the series obtained on combining them.*

Sol. Let $x_1, x_2 \ldots x_{n_1}$ and $y_1, y_2 \ldots y_{n_2}$ be the items of two series respectively.

Then, $$g_1 = \left\{x_1.x_2.\ldots.x_{n_1}\right\}^{1/n_1}$$

and $$g_2 = \left\{y_1.y_2.\ldots.y_{n_2}\right\}^{1/n_2}$$

Let g be the G.M. of the combined series.

Then, $$g = \left(x_1.x_2.\ldots\ldots.x_{n_1}.y_1.y_2.\ldots\ldots.y_{n_2}\right)^{\frac{1}{n_1+n_2}}$$

$$= \left\{(x_1.x_2.\ldots.x_{n_1})(y_1.y_2.\ldots.y_{n_2})\right\}^{\frac{1}{n_1+n_2}}$$

$$= \left\{g_1^{\,n_1}.g_2^{\,n_2}\right\}^{\frac{1}{n_1+n_2}}$$

$$= (g_1)^{\frac{n_1}{n_1+n_2}}(g_2)^{\frac{n_2}{n_1+n_2}}$$

Ex. 1-21. *If $x_1, x_2 \ldots x_n$ be non-zero positive numbers with A.M. 'A', G.M. 'G' and H.M. 'H', show that*

$$A \geq G \geq H.$$

Sol. $$A = \frac{x_1 + x_2 + \ldots + x_n}{n}, G = (x_1.x_2.\ldots.x_n)^{1/n}$$

and $$\frac{1}{H} = \frac{\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}\right)}{n}$$

From inequalities

$$\frac{x_1 + x_2 + \ldots + x_n}{n} \geq (x_1.x_2.\ldots.x_n)^{1/n}$$

$\therefore$ $$A \geq G.$$

Also $$\frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}{n} \geq \left(\frac{1}{x_1}\cdot\frac{1}{x_2}\ldots\frac{1}{x_n}\right)^{1/n}$$

$\therefore$ $\frac{1}{H} \geq \frac{1}{G}$ or $G \geq H$

$\therefore$ $A \geq G \geq H.$

Ex. 1-22. *A variate takes values a, ar, $ar^2 .... ar^{n-1}$ each with frequency unity. Show that A.M. 'A' is* $\dfrac{a(1-r^n)}{n(1-r)}$ *and G.M. 'G' is* $ar^{\frac{n-1}{2}}$ *and the H.M. 'H' is* $\dfrac{an(1-r)r^{n-1}}{1-r^n}$. *Prove that* $AH = G^2$. *Prove also that* $A \geq G \geq H$.

Sol. $A = \dfrac{a(1+r+r^2+....+r^{n-1})}{n} = \dfrac{a(1-r^n)}{n(1-r)}$

$$G = (a.ar.ar^2,...,ar^{n-1})^{1/n} = a\,r^{\frac{1+2+...+\overline{n-1}}{n}}$$

$$= ar^{\frac{n-1}{2}}$$

$$\frac{1}{H} = \frac{1}{na}\left\{1+\frac{1}{r}+\frac{1}{r^2}+....+\frac{1}{r^{n-1}}\right\} = \frac{1}{na}\cdot\frac{1-\dfrac{1}{r^n}}{1-\dfrac{1}{r}}$$

$$\therefore\ H = \frac{na(1-r)r^{n-1}}{1-r^n}$$

$$\therefore\ A.H. = a^2r^{n-1} = G^2.$$

***(iii)* Partition values :** *These are the values of the variate which divide the total frequency into a number of equal parts. Some important partition values are quartiles, deciles, percentiles, quintiles, etc. For a grouped dist these are given by*

$$Q_i = L + \left\{\frac{i\dfrac{N}{4} - c}{f}\right\} h \quad i = 1,2,3$$

Q_2 *is known as median.*

$$D_j = L + \left\{\frac{j\dfrac{N}{10} - c}{f}\right\}\cdot h \qquad j = 1, 2....9$$

$$P_k = L + \left\{\frac{k\dfrac{N}{100} - c}{f}\right\}\cdot h \qquad k = 1, 2....99$$

and l^{th} quintile $= L + \left\{\dfrac{l\dfrac{N}{5} - c}{f}\right\}\cdot h \qquad l = 1, 2, 3, 4$

where L = *Lower limit o*
h = *Width of the*
f = *Frequency o*
c = *Cumulative f*
which partiti
N = Total Freque

Ex. 1-23. *Find out the*
25, 15, 23, 40, 27,

Sol. Items arranged in a

If n is the number of ite

Median = size of $\left(\dfrac{n+1}{2}\right.$

= size of $\left\{\left(\dfrac{9+}{2}\right.\right.$

= 25.

Ex. 1-24. *From the dat*

Sol. Given figures arra

S.N.
1
2
3
4
5
6
7
8
9

Median = value of

= value of 9·

= value of 9

th frequency unity. Show

' is $\dfrac{an(1-r)r^{n-1}}{1-r^n}$. Prove

divide the total frequency
s are quartiles, deciles,

9

where L = *Lower limit of the class in which partition value lies.*
h = *Width of the class.*
f = *Frequency of the class.*
c = *Cumulative frequency upto and including the class preceding the class in which partition value lies.*
N = Total Frequency.

Ex. 1-23. *Find out the median of the following items :*
25, 15, 23, 40, 27, 25, 23, 25 *and* 20.

Sol. Items arranged in ascending order of magnitude :

S.N.	*Size of the items*
1	15
2	20
3	23
4	23
5	25
6	25
7	25
8	27
9	40

If n is the number of items,

$$\text{Median} = \text{size of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item}$$

$$= \text{size of } \left\{\left(\frac{9+1}{2}\right)^{\text{th}} = 5^{\text{th}}\right\} \text{ item}$$

$$= 25.$$

Ex. 1-24. *From the data of Ex. 1-3 find out the median and Quartiles.*

Sol. Given figures arranged in ascending order are .

S.N.	*Marks*	*S.N.*	*Marks*
1	11	10	22
2	11	11	23
3	12	12	32
4	15	13	32
5	15	14	33
6	17	15	35
7	18	16	35
8	20	17	38
9	21	18	41

$$\text{Median} = \text{value of } \left(\frac{18+1}{2}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } 9{\cdot}5^{\text{th}} \text{ item}$$

$$= \frac{\text{value of } 9^{\text{th}} \text{ item} + \text{value of } 10^{\text{th}} \text{ item}}{2}$$

$$= \frac{21+22}{2} = \frac{43}{2} = 21{\cdot}5$$

$$Q_1 = \text{value of } \left(\frac{18+1}{4}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } (4{\cdot}75)^{\text{th}} \text{ item}$$

$$= \text{value of } 4^{\text{th}} \text{ item} + \frac{3}{4} \text{ (value of } 5^{\text{th}} \text{ item—value of } 4^{\text{th}} \text{ item)}$$

$$= 15 + \frac{3}{4}(15-15) = 15$$

$$Q_3 = \text{value of } \frac{3(18+1)^{\text{th}}}{4} \text{ item}$$

$$= \text{value of } (14{\cdot}25)^{\text{th}} \text{ item}$$

$$= \text{value of } 14^{\text{th}} \text{ item} + \frac{1}{4} \text{ (value of } 15^{\text{th}} \text{ item—value of } 14^{\text{th}} \text{ item)}$$

$$= 33 + \frac{1}{4}(35-33)$$

$$= 33{\cdot}5.$$

Ex. 1-25. *From the data given below, find out the median and the two Quartiles :*

Wages in Rs. :	20	21	22	23	24	25	26	27	28
No. of workers :	8	10	11	16	20	25	15	9	6

Sol.

Wages in Rs.	*Frequency No. of workers*	*Cumulative Frequency*
20	8	8
21	10	18
22	11	29
23	16	45
24	20	65
25	25	90
26	15	105
27	9	114
28	6	120

$$\text{Median} = \text{value of } \left(\frac{120+1}{2}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } (60{\cdot}5)^{\text{th}} \text{ item}$$

$$= \frac{\text{value of } 60^{\text{th}} \text{ item} + \text{value of } 61^{\text{st}} \text{ item}}{2}$$

From the above table, there are 45 items upto 23 and 65 items upto 24.

∴ Value of item from 46^{th} to 65^{th} is 24.

∴ Value of 60^{th} and 61^{st} items each is 24.

$$\therefore \text{Median} = \frac{24+24}{2} = 24$$

$$Q_1 = \text{value of } \left(\frac{120+1}{4}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } (30{\cdot}25)^{\text{th}} \text{ item}$$

$$= \text{value of } 30^{\text{th}} \text{ item} + \frac{1}{4} \text{ (value of } 31^{\text{st}} \text{ item} - \text{value of } 30^{\text{th}} \text{ item)}$$

$$= 23 + \frac{1}{4}(23-23)$$

$$= 23$$

$$Q_3 = \text{value of } \frac{3}{4}(120+1)^{\text{th}} \text{ item}$$

$$= \text{value of } (90{\cdot}75)^{\text{th}} \text{ item}$$

$$= \text{value of } 90^{\text{th}} \text{ item} + \frac{3}{4} \text{ (value of } 91^{\text{st}} \text{ item} - \text{value of } 90^{\text{th}} \text{ item)}$$

$$= 25 + \frac{3}{4}(26-25)$$

$$= 25{\cdot}75.$$

Ex. 1-26. *Find out the median, quartiles, 3rd quintile, 5th octile, 7th decile from the following data :*

Monthly Rent in Rs.	*No. of families paying the Rent*
20— 40	6
40— 60	9
60— 80	11
80—100	14
100—120	20
120—140	15
140—160	10
160—180	8
180—200	7

Sol.

Class-Intervals	*Frequency*	*Cumulative Frequency*
20— 40	6	6
40— 60	9	15
60— 80	11	26
80—100	14	40
100—120	20	60
120—140	15	75
140—160	10	85
160—180	8	93
180—200	7	100

$$\text{The Median} = \text{value of} \left(\frac{100}{2}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } 50^{\text{th}} \text{ item}$$

which lies in 100—120.

Applying interpolation formula,

$$\text{Median} = 100 + \frac{(120-100)}{20}(50-40)$$

$$= 100 + 10 = 110$$

$$Q_1 = \text{value of} \left(\frac{100}{4}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } 25^{\text{th}} \text{ item}$$

which lies in (60—80)

$$\therefore \quad Q_1 = 60 + \frac{(80-60)}{11}\{25-15\}$$

$$= 60 + \frac{20}{11}(10)$$

$$= 60 + \frac{200}{11} = 60 + 18{\cdot}2 = 78{\cdot}2$$

$$Q_3 = \text{value of } \frac{3}{4}(100)^{\text{th}} \text{ item}$$

$$= \text{value of } 75^{\text{th}} \text{ item}$$

which lies in 120—140

$$\therefore\ Q_3 = 120 + \frac{140-120}{15}(75-60) = 140$$

$$3^{\text{rd}} \text{ quintile} = \text{value of } \frac{3}{5}(100)^{\text{th}} \text{ item}$$

$$= \text{value of } 60^{\text{th}} \text{ item}$$

which lies in 100—120

$$\therefore\ 3^{\text{rd}} \text{ quintile} = 100 + \frac{120-100}{20}(60-40) = 120.$$

$$5^{\text{th}} \text{ octile} = \text{value of } \frac{5}{8}(100)^{\text{th}} \text{ item}$$

$$= \text{value of } 62{\cdot}5$$

which lies in (120—140)

$$\therefore\ 5^{\text{th}} \text{ octile} = 120 + \frac{140-120}{15}(62{\cdot}5-60)$$

$$= 120 + \frac{4}{3}(2{\cdot}5) = 123{\cdot}3$$

7^{th} Decile = value of $\frac{7}{10}$

= value of 70

which lies in 120—140

$\therefore\ 7^{\text{th}}$ Decile = $120 + \frac{140}{1}$

= $120 + \frac{4}{3}$ (1

Ex. 1-27. *Find the median f*

Monthly Wages (in Rs.) 50—5

No. of Workers : 6

Sol.

Class Intervals
50— 55
55— 60
60— 65
65— 70
70— 75
75— 80
80—100

Median = value of $\left(\frac{111}{2}\right)^{\text{th}}$

= value of $55{\cdot}5^{\text{th}}$ ite

which lies in 65—70

$\therefore$ Median = $65 + \frac{70-6}{30}$

= $65 + \frac{1}{6}$ (17·

Ex. 1-28. *Find the Median*

Sol.

Class Intervals
100—200
200—300
300—400
400—500
500—600

Median = value of $\frac{100^{\text{th}}}{2}$ it

= value of 50^{th} item

which lies in 300—400

$$7^{th} \text{ Decile} = \text{value of } \frac{7}{10}(100)^{th} \text{ item}$$
$$= \text{value of } 70^{th} \text{ item}$$

which lies in 120—140

$$\therefore 7^{th} \text{ Decile} = 120 + \frac{140-120}{15}(70-60)$$

$$= 120 + \frac{4}{3}(10) = 133\cdot 3.$$

Ex. 1-27. *Find the median for the data :*

Monthly Wages (in Rs.)	50—55	55—60	60—65	65—70	70—75	75—80	80—100
No. of Workers :	6	10	22	30	16	12	15

Sol.

Class Intervals	*Frequency*	*Cumulative Frequency*
50— 55	6	6
55— 60	10	16
60— 65	22	38
65— 70	30	68
70— 75	16	84
75— 80	12	96
80—100	15	111

$$\text{Median} = \text{value of } \left(\frac{111}{2}\right)^{th} \text{ item}$$

$$= \text{value of } 55\cdot 5^{th} \text{ item}$$

which lies in 65—70

$$\therefore \text{Median} = 65 + \frac{70-65}{30}(55\cdot 5 - 38)$$

$$= 65 + \frac{1}{6}(17\cdot 5) = 67\cdot 92.$$

Ex. 1-28. *Find the Median and Quartiles from the data of Ex. 1-5.*

Sol.

Class Intervals	*Frequency*	*Cumulative Frequency*
100—200	15	15
200—300	18	33
300—400	30	63
400—500	20	83
500—600	17	100

$$\text{Median} = \text{value of } \frac{100^{th}}{2} \text{ item}$$

$$= \text{value of } 50^{th} \text{ item}$$

which lies in 300—400

950/-

$$\therefore \text{Median} = 300 + \frac{400-300}{30}(50-33)$$

$$= 300 + \frac{100}{30}(17) = 356{\cdot}67$$

$$Q_1 = \text{Value of } \frac{100^{\text{th}}}{4} \text{ item}$$

$$= \text{Value of } 25^{\text{th}} \text{ item}$$

which lies in 200—300.

$$\therefore \quad Q_1 = 200 + \frac{300-200}{18}(25-15)$$

$$= 200 + \frac{100}{18}(10) = 255{\cdot}55$$

$$Q_3 = \text{Value of } \left[\frac{3}{4}(100)\right]^{\text{th}} \text{ item}$$

$$= \text{Value of } (75)^{\text{th}} \text{ item}$$

which lies in 400—500

$$\therefore \quad Q_3 = 400 + \frac{500-400}{20}(75-63)$$

$$= 400 + 5(12) = 460$$

Ex. 1-29. *Compute the Median of data in Ex. 1-7.*

Sol.

Class Frequency	*Frequency*	*Cumulative Frequency*
100—119	8	8
120—139	10	18
140—159	16	34
160—179	15	49
180—199	10	59
200—239	8	67
240—259	3	70

$$\text{Median} = \text{Value of } \left(\frac{70}{2}\right)^{\text{th}} \text{ item}$$

$$= \text{Value of } 35^{\text{th}} \text{ item}$$

which lies in 160—179.

As class intervals are of inclusive type, the real limits of the group are 159·5 to 179·5.

$$\therefore \text{Median} = 159{\cdot}5 + \frac{179{\cdot}5 - 159{\cdot}5}{15}(35-34)$$

$$= 159{\cdot}5 + \frac{20}{15} = 160{\cdot}83$$

(*iv*) **Mode :** *It is that value o*
dist it is given by

$$\text{Mode} = L + \frac{f_m - f_1}{2f_m - f_1 - f_2}$$

$$\text{Mode} = L + \frac{f_2}{f_1 + f_2} h$$

Second Method is used whe
where L = *Lower limit of t*
f_m = *Frequency of m*
f_1 = *Frequency of th*
f_2 = *Frequency of th*

For a moderately skew distr
Mode = 3 Median – 2

Ex. 1-30. *From the data of*

Sol. Converting the data int

Marks	*Fr*
11	
12	
15	
17	
18	
20	
21	

In the table on next page, t
columns (2) and (3), then in three
(10) and in fives in columns (11)
column is indicated by putting a s
To find out the point of maxir
table below :

Col.	11	12	15	17	1
(1)	✓		✓		
(2)	✓	✓	✓	✓	
(3)		✓	✓		
(4)	✓	✓	✓		
(5)		✓	✓	✓	
(6)			✓	✓	✓
(7)	✓	✓	✓	✓	
(8)					

MATHEMATICAL STATISTICS

Cumulative Frequency
8
18
34
49
59
67
70

f the group are 159·5 to 179·5.

(*iv*) **Mode :** *It is that value of the variate for which frequency is maximum. For grouped dist it is given by*

$$\text{Mode} = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} h \qquad \text{(I Method)}$$

$$\text{Mode} = L + \frac{f_2}{f_1 + f_2} h \qquad \text{(II Method)}$$

Second Method is used where first fails,

where L = *Lower limit of the class in which mode lies i.e., modal class.*

f_m = *Frequency of modal class.*

f_1 = *Frequency of the class preceding the modal class.*

f_2 = *Frequency of the class following the modal class.*

For a moderately skew distribution mode is given by

Mode = 3 Median – 2 Mean

Ex. 1-30. *From the data of Ex. 1-3 find the mode :*

Sol. Converting the data into ordinary Frequency dist :

Marks	*Frequency*	*Marks*	*Frequency*
11	2	22	1
12	1	23	1
15	2	32	2
17	1	33	1
18	1	35	2
20	1	38	1
21	1	41	1

In the table on next page, the frequencies in column (1) are first added in two's in columns (2) and (3), then in three's in columns (4), (5), (6), in four's in columns (7), (8), (9), (10) and in fives in columns (11), (12), (13), (14), (15). The maximum frequency in each column is indicated by putting a sign '✓' above the figure.

To find out the point of maximum concentration the data can be arranged in the shape of table below :

Analysis Table

Col.	11	12	15	17	18	20	21	22	23	32	33	35	38	41
(1)	✓		✓							✓		✓		
(2)	✓	✓	✓	✓					✓	✓	✓	✓		
(3)		✓	✓							✓	✓	✓	✓	
(4)	✓	✓	✓							✓	✓	✓		
(5)		✓	✓	✓				✓	✓	✓	✓	✓	✓	
(6)			✓	✓	✓				✓	✓	✓	✓	✓	✓
(7)	✓	✓	✓	✓					✓	✓	✓	✓		
(8)										✓	✓	✓	✓	

(*Contd.*)

Location of Mode by Grouping

Marks	Frequency (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
11	2 ✓	3 ✓		5 ✓			6 ✓				7 ✓				
12	1		3 ✓		4 ✓			5				6 ✓			
15	2 ✓	3 ✓				4 ✓			5 ✓				6		
17	1		2	3						4				5	
18	1	2			3		4								5
20	1		2			3		4			6				
21	1	2		3					5 ✓			6 ✓			
22	1		2		4 ✓		6 ✓			5 ✓			7 ✓		
23	1	3 ✓				4 ✓								7 ✓	
32	2 ✓		3 ✓	5 ✓				6 ✓							7 ✓
33	1	3 ✓			4 ✓				5 ✓						
35	2 ✓		3 ✓			4 ✓									
38	1	2													
41	1														

(9)			✓	✓	
(10)					
(11)	✓	✓	✓	✓	
(12)		✓	✓	✓	
(13)					
(14)					
(15)					
	5	7	10	7	

In this table marks are taker
vertical. Since according to colu
row (1) the signs '✓' are put und
(2), mode should be either 11 or
'✓' are put under 11, 12, *etc.*

In this way the whole table i

Since value 32 occurs the la

Ex. 1-31. *Following is the d*
from a district. Calculate the mo

Central size
(in a
1(
2(
3(
4(
5(
6(
7(

Sol. Since central size increa
of size 10. Hence various class-ii

Model class is 35—45.

$\therefore$ Mo

which lies outside the class-interv

$\therefore$ The formula fails.

In such cases we use the seco

Mo

By this formula,

Mo

(9)			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(10)								✓	✓	✓	✓			
(11)	✓	✓	✓	✓	✓									
(12)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
(13)								✓	✓	✓	✓	✓		
(14)									✓	✓	✓	✓	✓	
(15)										✓	✓	✓	✓	✓
	5	7	10	7	4	2	2	5	9	14	13	12	7	3

In this table marks are taken along the horizontal and columns (1), (2), *etc.*, along the vertical. Since according to column (1), mode should be either 11 or 15 or 32 or 35, in the row (1) the signs '✓' are put under 11, 15, 32 and 35. Similarly, since according to column (2), mode should be either 11 or 12 or 15 or 17 or 23 or 32 or 33 or 35, in row (2) the sign '✓' are put under 11, 12, *etc.*

In this way the whole table is completed.

Since value 32 occurs the largest number of times, mode is 32.

Ex. 1-31. *Following is the distribution of the size of certain farms selected at random from a district. Calculate the mode of the distribution :*

Central size of the farms (in acre)	*No. of farms*
10	8
20	12
30	17
40	29
50	31
60	5
70	3

Sol. Since central size increases by 10 throughout the data, each class interval must be of size 10. Hence various class-intervals are 5—15, 15—25 *etc.* (*See Table on page 26*).

Model class is 35—45.

$$\therefore \qquad \text{Model} = 35 + \frac{29 - 17}{58 - 17 - 31}(10)$$

$$= 35 + \frac{120}{10} = 47$$

which lies outside the class-interval 35—45.

$\therefore$ The formula fails.

In such cases we use the second formula, *i.e.*,

$$\text{Mode} = L + \frac{f_2}{f_1 + f_2} h$$

By this formula,

$$\text{Mode} = 35 + \frac{31}{17 + 31}(10) = 41{\cdot}46.$$

Class intervals	Frequency (1)	(2)	(3)	(4)	(5)	(6)	Columns (1)	(2)	(3)	(4)	(5)	(6)	
5—15	8	20		37									
15—25	12		29		✓ 58						✓		1
25—35	17	✓ 46				✓ 77		✓			✓	✓	3
35—45	29		✓ 60	✓ 65				✓	✓	✓	✓	✓	5
45—55	✓ 31	36			39		✓		✓	✓		✓	4
55—65	5		8							✓			1
65—75	3												

Ex. 1-32. *Find out mode for the data of Ex. 1-5*

Income	Frequency (1)	(2)	(3)	(4)	(5)	(6)	Columns (1)	(2)	(3)	(4)	(5)	(6)	
100—200	15	33		✓						✓			1

55—65	5	36		39		✓		1
65—75	3		8					

Ex. 1-32. *Find out mode for the data of Ex. 1-5*

Income	*Frequency*						*Columns*						
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)	
100—200	15	33		✓ 63						✓			1
200—300	18		✓ 48		✓ 68				✓	✓	✓		3
300—400	✓ 30	✓ 50				✓ 67	✓	✓	✓	✓	✓	✓	6
400—500	20		37					✓			✓	✓	3
500—600	17											✓	1

$\therefore$ Modal class is 300—400

$$\therefore \text{Mode} = 300 + \frac{30-18}{60-18-20}(100) = 354\cdot 55$$

Ex. 1-33. *Find the mode from the data of Ex. 1-7*

Salary	Frequency (1)	Frequency (2)	Frequency (3)	Frequency (4)	Frequency (5)	Frequency (6)	Columns (1)	Columns (2)	Columns (3)	Columns (4)	Columns (5)	Columns (6)	
100—119	8	18		✓ 34						✓			1
120—139	10		✓ 26		✓ 41				✓	✓	✓		3
140—159	✓ 16	✓ 31				✓ 41	✓	✓	✓	✓	✓	✓	6
160—179	15		25	33				✓			✓	✓	3
180—199	10	18			21							✓	1
200—239	8		11										
240—259	3												

∴ Modal class is 140—15

i.e., 139·5—159·5

∴ M

Ex. 1-34. *If the mode and*
16 inches and 20·2 inches. Con
Sol. For a moderately asy
mode is

M

∴ Me

Ex. 1-35. *If the mode and*
16″ and 15·6″, what would be

Sol. M

M

Me

Ex. 1-36. *The dist.* $x_1, x_2 ..$
$X_1, X_2 ... X_n$ *with the same co*
'a' and 'b' are constants. Show
of the first dist by the same tra

Sol. By given,

∴

∴

$\therefore$ Modal class is 140—159

i.e., 139·5—159·5

$$\therefore \qquad \text{Mode} = 139\cdot5 + \frac{16-10}{32-10-15}(20)$$

$$= 139\cdot5 + \frac{(6)(20)}{7} = 156\cdot64.$$

Ex. 1-34. *If the mode and mean of a moderately asymmetrical series are respectively 16 inches and 20·2 inches. Compute the most probable median.*

Sol. For a moderately asymmetrical series, the relation connecting mean, median and mode is

$$\text{Mode} = \text{Mean} - 3\,(\text{Mean} - \text{Median})$$

$$= 3\ \text{Median} - 2\ \text{Mean}$$

$$\therefore \qquad \text{Median} = \frac{1}{3}\ (\text{Mode} + 2\ \text{Mean})$$

$$= \frac{1}{3}(16 + 40\cdot4)$$

$$= \frac{56\cdot4}{3} = 18\cdot8 \ \text{inches.}$$

Ex. 1-35. *If the mode and mean of a moderately asymmetrical series are, respectively 16″ and 15·6″, what would be its most probable median.*

Sol.

$$\text{Mean} = 15\cdot6''$$

$$\text{Mode} = 16''$$

$$\text{Median} = \frac{1}{3}\ (\text{mode} + 2\ \text{mean})$$

$$= \frac{1}{3}(16 + 31\cdot2)$$

$$= \frac{47\cdot2}{3} = 15\cdot73$$

Ex. 1-36. *The dist.* $x_1, x_2 \ldots x_n$ *with frequencies* $f_1, f_2 \ldots f_n$ *is transformed into the dist.* $X_1, X_2 \ldots X_n$ *with the same corresponding frequencies by the relation* $X = ax + b$ *where 'a' and 'b' are constants. Show that the mean, median and mode are given in terms of those of the first dist by the same transformation.*

Sol. By given,

$$X = ax + b$$

$$\therefore \qquad X_i = ax_i + b$$

$$\therefore \qquad \overline{X} = \frac{1}{N}\sum_{i=1}^{n} f_i X_i = \frac{1}{N}\sum_{i=1}^{n} f_i (ax_i + b)$$

$$= a \cdot \frac{1}{N} \sum_{i=1}^{n} f_i x_i + b \cdot \frac{1}{N} \sum_{i=1}^{n} f_i = a\bar{x} + b$$

where $\overline{X}$ and $\bar{x}$ are A.Ms.

Since for median the value of the variable corresponding to the middle item is to be obtained and this middle item remains middle item in the transformed dist, median of the transformed dist is given by same transformation.

Similar argument holds for mode.

Ex. 1-37. *From the following data find the missing frequency :*

No. of Tablets :	4 — 8	— 12	— 16	— 20	— 24	— 28	— 32	— 36	— 40
No. of persons cured :	11	13	16	14	—	9	17	6	4

The average number of tablets to cure fever was 19·9.

Sol. Let the missing frequency be 'a'.

Class-Interval	*Frequency (f)*	*Mid-point (x)*	*xf*
4—8	11	6	66
8—12	13	10	130
12—16	16	14	224
16—20	14	18	252
20—24	a	22	$22a$
24—28	9	26	234
28—32	17	30	510
32—36	6	34	204
36—40	4	38	152
	$90 + a$		$22a + 1772$

Now A.M. of tablets for all the persons = 19·9.

$\therefore$ Total number of tablets for all the persons

$$= 19 \cdot 9\,(90 + a)$$
$$= 19 \cdot 9\, a + 1791$$

$\therefore$ $19 \cdot 9a + 1791 = 22a + 1772$

or $2 \cdot 1a = 19$

or $a = 9 \cdot 0.$

Ex. 1-38. *The following table gives the marks obtained by 30 students of a class in certain paper :*

Digits (Division) of class intervals

Marks	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Total*
30—39	2	1	2	2	—	1	—	1	1	—	10
40—49	—	1	—	2	—	—	4	—	—	—	7
50—59	—	—	—	—	3	1	—	3	—	1	8
60—69	1	1	—	—	1	—	1	—	—	1	5

Calculate the mean and median of the series : (a) by using only the total of class-intervals, (b) by using the entire data.

Sol. The data indicates that 4 students get 46 marks, 3 students get 54 marks, 3 students get 57 marks and so on.

$f_i = a\bar{x} + b$

to the middle item is to be
sformed dist, median of the

ncy :
—28—32—36—40
17 6 4

xf
66
130
224
252
22a
234
510
204
152
22a + 1772

by 30 students of a class in

7	8	9	Total
1	1	—	10
—	—	—	7
3	—	1	8
—	—	1	5

ing only the total of class-

nts get 54 marks, 3 students

Calculation of Mean

Marks	Total marks for each student *0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Total marks*	*Total Freq. (f)*	*Mid point (x)*	*xf*	*C Freq.*
30—39	60	31	64	66	—	35	—	37	38	—	331	10	34·5	345·0	10
40—49	—	41	—	86	—	—	184	—	—	—	311	7	44·5	311·5	17
50—59	—	—	—	—	162	55	—	171	—	59	447	8	54·5	436·0	25
60—69	60	61	—	—	64	—	66	—	—	69	320	5	64·5	322·5	30
Total	120	133	64	152	226	90	250	208	38	128	1409	30		1415	

Mean : (*a*) Using totals of classes

$$= \frac{\sum fx}{\sum f} = \frac{1415}{30} = 47\cdot2$$

$$= 47 \text{ (approx.)}$$

(*b*) Using entire data.

$$= \frac{\sum x}{n} = \frac{1409}{30} = 46 \cdot 97$$

$$= 47 \text{ (approx.)}$$

Median

(*a*) Using only totals of class intervals.

Median has $\frac{30}{2} = 15$ items below it *i.e.*, it lies in 40—49 *i.e.*, 39·5—49·5 (taking real limits).

$$\therefore \qquad \text{Median} = 39 \cdot 5 + \frac{10}{7}(15 - 10)$$

$$= 39 \cdot 5 + \frac{50}{7} = 46 \cdot 64.$$

(*b*) Using entire data.

$$\text{Median} = \text{value of} \left(\frac{30+1}{2}\right)\text{th item}$$

$$= \text{value of } 15 \cdot 5\text{th item}$$

$$= \frac{\text{value of 15th item} + \text{value of 16th item}}{2}$$

From Cumulative frequencies it is clear that 15th and 16th items lie in 40—49. There are 10 items upto 39.

So counting in the group 40—49 the various items we see that

15th item = 46 = 16th item

$\therefore$ Median = 46.

Ex. 1-39. *Show that in finding the arithmetic mean of a set of readings on a thermometer, it does not matter whether we measure the temperature in centigrade or Fahrenheit degrees, but that in finding the G.M. it does matter.*

Sol. Let a set of N thermometric readings in Centigrade degrees be $C_1, C_2 \ldots\ldots\ldots\ldots C_N$ and the corresponding readings in Fahrenheit degrees be $F_1, F_2 \ldots\ldots\ldots\ldots F_N$.

The relation between Centigrade and Fahrenheit readings is

$$F = 32 + \frac{9}{5}C.$$

where C corresponds to Centigrade readings and F to Fahrenheit readings

$$\therefore \qquad F_r = 32 + \frac{9}{5}C_r \qquad r = 1, 2 \ldots. N.$$

Now the A.M. of the N readings in Centigrade degrees

$$= \frac{C_1 + C_2 + \ldots. + C_N}{N} = \overline{C} \text{ (say)}$$

and the same in Fahrenheit degrees

$$= \frac{F_1 + F_2 + \ldots. + F_N}{N} = \overline{F} \textit{ (say)}$$

$$\therefore \qquad \overline{F} = \frac{1}{N}(F_1 + F_2 + \ldots + F_N)$$

$$= \frac{1}{N}\left\{\left(32 + \frac{9}{5}C_1\right) + \left(32 + \frac{9}{5}C_2\right) + \ldots\ldots\ldots + \left(32 + \frac{9}{5}C_N\right)\right\}$$

$$= 32 + \frac{9}{5}\left(\frac{C_1 + C_2 + \ldots + C_N}{N}\right) = 32 + \frac{9}{5}\overline{C}.$$

= Fahrenheit equivalent of $\overline{C}$ (A.M. of Centigrade readings).

G.M. of the readings in Fahrenheit

$$= (F_1, F_2 \ldots F_N)^{1/N}$$

$$= \left\{\left(32 + \frac{9}{5}C_1\right)\left(32 + \frac{9}{5}C_2\right)\ldots\left(32 + \frac{9}{5}C_N\right)\right\}^{1/N} \qquad \ldots(1)$$

G.M. of the readings in Centigrade

$$= (C_1 . C_2 \ldots C_N)^{1/N}$$

$\therefore$ Fahrenheit equivalent of the geometric mean of the readings in Centigrade

$$= 32 + \frac{9}{5}(C_1 C_2 \ldots C_N)^{1/N} \qquad \ldots(2)$$

But (1) and (2) are not same.

$\therefore$ Fahrenheit equivalent of the G.M. of the Centigrade readings is not the same as the G.M. of the Farenheit readings.

$\therefore$ The given statement follows.

Weighted Average. *If* $w_1 . w_2 \ldots\ldots w_n$ *be the weights of values* $x_1, x_2 \ldots x_n$, then

$$\textit{Weighted arithmetic mean} = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n}$$

$$\textit{Weighted geometric mean} = \textit{Antilog}\left\{\frac{w_1 \log x_1 + \ldots + w_n \log x_n}{w_1 + \ldots + w_n}\right\}$$

$$\textit{and Weighted harmonic mean} = \textit{Reciprocal}\left\{\frac{w_1 \dfrac{1}{x_1} + \ldots + w_n \dfrac{1}{x_n}}{w_1 + \ldots + w_n}\right\}.$$

Ex. 1·40. *Show that the weighted arithmetic mean of first n natural numbers when weights are equal to the corresponding numbers is* $\dfrac{2n+1}{3}$.

Sol. $$\text{Weighted A.M.} = \frac{1 \cdot 1 + 2 \cdot 2 + \ldots + n.n}{1 + 2 + \ldots + n}$$

$$= \frac{1^2+2^2+....+n^2}{\frac{n(n+1)}{2}}$$

$$= \frac{n(n+1)(2n+1)}{6} \cdot \frac{2}{n(n+1)}$$

$$= \frac{2n+1}{3}.$$

Ex. 1-41. *From the following results of two colleges A and B find out which of the two is better :*

	A		*B*	
Exam.	*Appeared*	*Passed*	*Appeared*	*Passed*
M.A.	100	90	240	200
M.Sc.	60	45	200	160
B.A.	120	75	160	100
B.Sc.	200	150	200	140
Total	480	360	800	600

Sol. Since the number of students appearing for M.A., M.Sc., B.A. and B.Sc. widely differ, simple arithmetic average of pass percentages of the college will not give the correct idea of pass percentage of a college for all the examinations taken together. So we take the weighted average of pass percentages, weights being the number of students appeared for each examination.

For college A,

Pass percentage for M.A. $= 90\%$

Pass percentage for M.Sc. $= \frac{45}{60} \times 100 = 75\%$

Pass percentage for B.A. $= \frac{75}{120} \times 100 = 62 \cdot 5\%$

Pass percentage for B.Sc. $= \frac{150}{200} \times 100 = 75\%.$

For college B,

Pass percentage for M.A. $= \frac{200}{240} \times 100 = \frac{250}{3}\%$

Pass percentage for M.Sc. $= \frac{160}{200} \times 100 = 80\%$

Pass percentage for B.A. $= \frac{100}{160} \times 100 = 62 \cdot 5\%$

Pass percentage for B.Sc. $= \frac{140}{200} \times 100 = 70\%$

	A
(x) *Pass %*	*No. of student* *appeared* *(weight) w*
90	100
75	60
62·5	120
75	200
	480

∴ Pass percentage of col

Pass percentage of col

Since pass percentages fc

Ex. 1-42. *Find the weigh*

Group
Fuod
Clothing
Fuel and li
House Ren
Miscellane

Sol.

Group	*Index*
Food	12
Clothing	13
Fuel and light	14
House Rent	17
Miscellaneous	18

∴ Weighted geometric n

∴

Ex. 1-43. *A train starts* *average speeds of 12, 16, 2* *19·2 k.m. per hour and not 2.*

A			*B*		
(x) *Pass %*	*No. of students appeared (weight) w*	*xw*	*(x)* *Pass %*	*No. of students appeared (weight) w*	*xw*
90	100	9000	$\frac{250}{3}$	240	20000
75	60	4500	80	200	16000
62·5	120	7500	62·5	160	10000
75	200	15000	70	200	14000
	480	36000		800	60000

$\therefore$ Pass percentage of college $A = \dfrac{36000}{480} = 75\%$

Pass percentage of college $B = \dfrac{60000}{800} = 75\%$

Since pass percentages for two colleges A and B are same, none is better than the other.

Ex. 1·42. *Find the weighted geometric mean from the following data :*

Group	*Index No.*	*Weight*
Food	125	7
Clothing	133	5
Fuel and light	141	4
House Rent	173	1
Miscellaneous	182	3

Sol.

Group	*Index No. x*	*Weight (w)*	$\log_{10} x$	$(w \log_{10} x)$
Food	125	7	2·0969	14·6783
Clothing	133	5	2·1239	10·6195
Fuel and light	141	4	2·1492	8·5968
House Rent	173	1	2·2380	2·2380
Miscellaneous	182	3	2·2601	6·7803
		20		42·9129

$\therefore$ Weighted geometric mean is given by

$$\log_{10} G = \frac{42\cdot9129}{20} = 2\cdot1457$$

$$\therefore \qquad G = 139\cdot8 = 140.$$

Ex. 1·43. *A train starts from rest and travels successive quarters of a kilometre at average speeds of 12, 16, 24, 48 k m per hour. The average speed over the whole k m is 19·2 k.m. per hour and not 25 k.m. per hour. Explain.*

Sol. Here average speeds for each quarter of a k.m. are given. To find out the average speed over the total distance, first the total time taken by the train is to be calculated by dividing distances by average speeds and then the total distance is to be divided by the total time. This procedure is equivalent to finding the weighted harmonic mean of average speeds weights being the respective distances.

Here distances travelled in four cases are same each being equal to $\frac{1}{4}$ k.m. So here equal weighted or simple harmonic mean is the appropriate method of averaging.

$\therefore$ Average speed over the whole mile

$$= \frac{1}{\frac{1}{4}\cdot\frac{1}{12}+\frac{1}{4}\cdot\frac{1}{16}+\frac{1}{4}\cdot\frac{1}{24}+\frac{1}{4}\cdot\frac{1}{48}}$$

$$= \frac{(4)(48)}{4+3+2+1} = 19\cdot 2.$$

$\therefore$ Average speed over the whole k.m. is 19·2 *k.m.p.h.* and not 25 *k.m.p.h.* which is simply the A.M. or weighted (weights being $\frac{1}{4}$ each) A.M. of four average speeds 12, 16, 24 and 48.

Ex. 1-44. *A cyclist covers his first three k.m. at an average speed of 8 k.m.p.h., another two k m at 3 k.m.p.h. and the last two k m at 3 k.m.p.h. Find his average speed for the entire journey.*

Sol. The average speed for the entire journey is the weighted harmonic mean of the speeds with distances as weights.

$\therefore$ The average speed for the entire journey

$$= \frac{3 \div 2 + 2}{3\cdot\frac{1}{8}+2\cdot\frac{1}{3}+2\cdot\frac{1}{3}} = \frac{7}{\frac{3}{8}+\frac{4}{3}}$$

$$= \frac{(7)(24)}{9+32} = \frac{168}{41} = 4\cdot 1\ k.m.p.h.$$

Ex. 1-45. *Mr. X travels from A to B at an average speed of 30 k.m.p.h. and returns from B to A at an average speed of 60 k m per hour. Find the average speed of Mr. X for the entire trip.*

Sol. Let x be the distance between A and B

Then average speed for the entire trip

$$= \frac{2x}{x\cdot\frac{1}{30}+x\cdot\frac{1}{60}} = \frac{(2)(60)}{2+1}$$

$$= 40.$$

$\therefore$ Average speed for the entire trip

$$= 40\ k.m.p.h.$$

Ex. 1-46. *An aeroplane flies round a square the sides of which measure 100 km each. The aeroplane covers at a speed of 100 km per hour the first side, at 200 k.m.p.h. the second side, at 300 k.m.p.h. the third side and 400 k.m.p.h. the fourth side. What is the average speed of the aeroplane around the square?*

Sol. Average spe

Ex. 1-47. *You take a trip which*
of 60 k.m.p.h., 3000 k.m. by boat a
350 k.m.p.h. and finally 15 k.m. by
entire distance (4315 k.m.) ?

Sol. The average speed

Ex. 1-48. *A man travels 50 km a*
k.m.p.h. What is his average speed f

Sol. Average speed for the whol

Ex. 1-49. *The price of a commodi*
to 2000 and 77% from 2000 to 2001.
26% and not 30%. Explain this staten

Sol. Let the price of the commod
beginning of 1999

Since the price from 1999 to 20

$$2000 = \frac{108}{100}\frac{105}{100}x.$$

given. To find out the average
he train is to be calculated by
nce is to be divided by the total
rmonic mean of average speeds

eing equal to $\frac{1}{4}$ k.m. So here

method of averaging.

$$\cdot\frac{1}{24}+\frac{1}{4}\cdot\frac{1}{48}$$

and not 25 *k.m.p.h.* which is

of four average speeds 12, 16,

ge speed of 8 k.m.p.h., another
is average speed for the entire

eighted harmonic mean of the

$$\frac{7}{\frac{3}{8}+\frac{4}{3}}$$

k.m.p.h.

f 30 k.m.p.h. and returns from
ge speed of Mr. X for the entire

$$\frac{(60)}{+1}$$

which measure 100 km each.
de, at 200 k.m.p.h. the second
rth side. What is the average

Sol.

$$\text{Average speed} = \frac{400}{100\left(\frac{1}{100}+\frac{1}{200}+\frac{1}{300}+\frac{1}{400}\right)}$$

$$= \frac{4(1200)}{12+6+4+3}$$

$$= \frac{4800}{25} = 192\ k.m.p.h.$$

Ex. 1-47. *You take a trip which entails travelling 900 k.m. by train at an average speed of 60 k.m.p.h., 3000 k.m. by boat at an average speed of 25 k.m.p.h., 400 k.m. by plane at 350 k.m.p.h. and finally 15 k.m. by taxi at 25 k.m.p.h. What is your average speed for the entire distance (4315 k.m.) ?*

Sol. The average speed

$$= \frac{900+3000+400+15}{900\cdot\frac{1}{60}+3000\cdot\frac{1}{25}+400\cdot\frac{1}{350}+15\cdot\frac{1}{25}}$$

$$= \frac{4315}{15+120+\frac{8}{7}+\frac{3}{5}}$$

$$= \frac{(4315)(35)}{525+4200+40+21}$$

$$= \frac{(4315)(35)}{4786} = \frac{151025}{4786} = 31\cdot 56\ k.m.p.h.$$

Ex. 1-48. *A man travels 50 km at a speed of 20 k.m.p.h. and then returns at speed of 30 k.m.p.h. What is his average speed for the whole journey ?*

Sol. Average speed for the whole journey

$$= \frac{50+50}{\left(50\cdot\frac{1}{20}\right)+\left(\frac{50}{30}\right)}$$

$$= \frac{(2)(60)}{3+2} = 24\ k.m.p.h.$$

Ex. 1-49. *The price of a commodity increased by 5% from 1998 to 1999, 8% from 1999 to 2000 and 77% from 2000 to 2001. The average increase from 1998 to 2001 is quoted as 26% and not 30%. Explain this statement and verify the arithmetic.*

Sol. Let the price of the commodity in the beginning of 1998 be x. Then price in the beginning of 1999

$$= \left(\frac{105}{100}x\right)$$

Since the price from 1999 to 2000 increases by 8%, the price in the beginning of

$$2000 = \frac{108}{100}\frac{105}{100}x.$$

Similarly, the price in the beginning of 2001

$$= \left(\frac{108}{100}\right)\left(\frac{105}{100}\right) x \cdot \left(\frac{177}{100}\right).$$

$\therefore$ The price at the end of 2000 or in the beginning of 2001

$$= \frac{(105)\,(108)\,(177)}{(100)^3}\, x$$

Let r be the average rate of increase. Then

$$\frac{(105)\,(108)\,(177)}{(100)^3}\, x = x\,(1+r)^3$$

or

$$1+r = \frac{1}{100}\sqrt[3]{(105)\,(108)\,(177)}$$

$$\therefore \quad \log_{10}(1+r) = \frac{1}{3}(-6 + 2\cdot 0212 + 2\cdot 0334 + 2\cdot 2480)$$

$$= \frac{1}{3}(0\cdot 3026) = 0\cdot 1009$$

$$\therefore \quad 1+r = 1\cdot 262$$

$$\therefore \quad r = 0\cdot 262 \text{ or } 26\%.$$

$\therefore$ Average price rise was 26%.

The A.M. of the rise in price is

$$\frac{5+8+77}{3} = 30\%.$$

If this be the rise in price in each year, the price at the end of 2000 would be $\left(\frac{130}{100}\right)^3 x$ which is much higher than the value obtained from the given data *i.e.*, $x\left(\frac{105}{100}\right)\left(\frac{108}{100}\right)\left(\frac{177}{100}\right)$.

But if the rate of rise in price is taken to be 26%, the price at the end of 2000 would be $x\left(\frac{126}{100}\right)^3$ which is nearly equal to the value obtained from the given data.

$\therefore$ Average price rise was 26% and not 30%.

Ex. 1-50. *Find the average rate of increase in population which in the first decade had increased 20%, in the next 30% and in the third 45%.*

Sol. Let x be the population in the beginning. Then the population at the end of first decade $= \left(\frac{120}{100}\right) x$.

Since the population in the next decade increases by 30%, the population at the end of second decade $= \left(\frac{130}{100}\right)\left(\frac{120}{100}\right) x$.

Similarly, the population at the end of third decade

$$= \left(\frac{145}{100}\right)\left(\frac{130}{100}\right)\left(\frac{120}{100}\right) x.$$

Let r be the average rate of increase per year.

Then

$$\frac{(145)\,(130)\,(120)}{(100)^3}\, x = x(1+r)^{30}$$

$$\therefore \quad \log_{10}(1+r) = \frac{1}{30}\{-6 + 2\cdot1614 + 2\cdot1139 + 2\cdot0792\}$$

$$= \frac{1}{30}\{0\cdot3545\} = 0\cdot0118$$

$$\therefore \quad 1 + r = 1\cdot028$$

$$\therefore \quad r = 0\cdot028 \text{ or } 2\cdot8\% = 3\%.$$

Ex. 1-51. *A machine is assumed to depreciate 40% in value in the first year, 25% in the second year and 10% per annum for the next three years, each percentage being calculated on the diminishing value. What is the average percentage depreciation for the five years ?*

Sol. Let x be the value of the machine in the beginning.

Then price at the end of first year

$$= \left(\frac{60}{100}\right) x.$$

The price at the end of 2nd year

$$= \left(\frac{75}{100}\right)\left(\frac{60}{100}\right) x.$$

The price at the end of fifth year

$$= \left(\frac{75}{100}\right)\left(\frac{60}{100}\right)\left(\frac{90}{100}\right)^3 x$$

Let r be the average rate of depreciation per year.

$$\text{Then} \quad \frac{(75)\,(60)\,(90)^3}{(100)^5}\, x = x(1-r)^5$$

$$\therefore \quad \log_{10}(1-r) = \frac{1}{5}\{-10 + 1\cdot8751 + 1\cdot7782 + 3(1\cdot9542)\}$$

$$= \frac{1}{5}\{-10 + 1\cdot8751 + 1\cdot7782 + 5\cdot8626\}$$

$$= \bar{1}\cdot9032$$

$$\therefore \quad (1-r) = 0\cdot8002$$

$$\Rightarrow \quad r = 19\cdot98\%$$

$$= 20\%.$$

Ex. 1-52. *The age distribution of the members of a certain children's club is as follows :*

Age on last birthday (in yrs.)	*Frequency*
4	5
5	9
6	18
7	35
8	42
9	32
10	15
11	7
12	3

There is a member A s.t. there are twice as many members older than A as there are younger than A. Estimate his age (in years upto two places of decimals).

Sol. Since ages on last birthday are given,

No. of persons who are in 4—5 group = 5

No. of persons who are in 5—6 group = 9

and so on.

Age	*Frequency (f)*	*Cumulative Freq.*
4—5	5	5
5—6	9	14
6—7	18	32
7—8	35	67
8—9	42	109
9—10	32	141
10—11	15	156
11—12	7	163
12—13	3	166

$$\text{Age of } A = \text{size of } \left(\frac{166}{3}\right)^{\text{th}} \text{ item}$$

$$= \text{size of } \left(55\frac{1}{3}\right)^{\text{th}} \text{ item}$$

which lies in 7—8

$$\therefore \qquad \text{Age of } A = 7 + \frac{(8-7)}{35}\left(55\frac{1}{3} - 32\right)$$

$$= 7 + \frac{1}{35}\left(23\frac{1}{3}\right)$$

$$= 7 + \frac{70}{105}$$

$$= 7{\cdot}67.$$

Ex. 1-53. *An incomplete f*

Variate : 10—20 20—30

Freq. : 12 30

Given that median value i

Sol. Median = 46 lies in 4

$\therefore$

where x is the freq. of the class

$\therefore \; x = 33{\cdot}5 \simeq 34.$

$\therefore$ Freq. of class 50—60 =

=

Find arithmetic means of foll

1. Gold output (in millions o

94	95	96
78	82	83

2. | | | |
|---|---|---|
| $x : 0$ | 1 | 2 |
| $f : 1$ | 9 | 26 |

where x denotes the numb tossed 256 times.

3. | *Age Group* | *No. of pe* |
|---|---|
| 25—30 | 1 |
| 30—35 | 2 |
| 35—40 | 4 |
| 40—45 | 10 |
| 45—50 | 21 |

4. Find A.M. of data in *Ex.* 1-

5. | *Wts (in lbs.)* | *Freq.* |
|---|---|
| 90—100 | 10 |
| 100—110 | 37 |
| 110—120 | 65 |
| 120—130 | 80 |

6. | *Wage (in Rs.)* | *No. of E* |
|---|---|
| 50—55 | 2 |
| 45—50 | 3 |
| 40—45 | 4 |
| 35—40 | 4 |

children's club is as follows :

Frequency
5
9
18
35
42
32
15
7
3

rs older than *A* as there are
decimals).

nulative Freq.
5
14
32
67
109
141
156
163
166

Ex. 1-53. *An incomplete freq. dist. is given below :*

Variate :	10—20	20—30	30—40	40—50	50—60	60—70	70—80	Total
Freq. :	12	30	?	65	?	25	18	229

Given that median value is 46, find the missing frequencies using the median formula.

Sol. Median = 46 lies in 40—50 class

$$\therefore \qquad 46 = 40 + 10\left\{\frac{\frac{229}{2} - (x+42)}{65}\right\}$$

where x is the freq. of the class 30—40.

$\therefore \ x = 33\cdot5 \simeq 34.$

$\therefore$ Freq. of class 50—60 = 229 — (Sum of remaining frequencies)
$= 229 - 184 = 45$

EXERCISES

Find arithmetic means of following datas :

1. Gold output (in millions of pounds) for different years.

94	95	96	93	87	79	73	69	68	67
78	82	83	89	95	103	108	117	130	97

(Ans. 90·15)

2.

x :	0	1	2	3	4	5	6	7	8
f :	1	9	26	59	72	52	29	7	1

where x denotes the number of heads and f their frequencies when eight coins are tossed 256 times.

(Ans. 3·97)

3.

Age Group	*No. of persons*	*Age Group*	*No. of persons*
25—30	1	50—55	53
30—35	2	55—60	126
35—40	4	60—65	163
40—45	10	65—70	35
45—50	21	70—75	6
		75—80	1

(Ans. 58·66)

4. Find A.M. of data in *Ex.* 1-26. **(Ans.** 110)

5.

Wts (in lbs.)	*Freq.*	*Wts (in lbs.)*	*Freq.*
90—100	10	130—140	51
100—110	37	140—150	35
110—120	65	150—160	18
120—130	80	160—170	4

(Ans. 125·73)

6.

Wage (in Rs.)	*No. of Employees*	*Wage (in Rs.)*	*No. of Employees*
50—55	250	30—35	800
45—50	300	25—30	1100
40—45	400	20—25	1700
35—40	450		

(Ans. 31.15)

7.

No. of days absent	*No. of students*	*No. of days absent*	*No. of students*
Less than 5	29	Less than 30	487
Less than 10	124	Less than 35	493
Less than 15	349	Less than 40	497
Less than 20	442	Less than 45	500
Less than 25	478		

(Ans. 13·51)

8. Find out the median from the following data :

Age group (in years)	15—20	20—25	25—30	30—35	35—40
No. of men	5	9	82	58	49
Age group (in years)	40—45	45—50	50—55		
No. of men	28	6	3		

(Ans. 32·07)

9. From the following data find out the median and the quartiles :

Marks	*No. of students*	*Marks*	*No. of students*
0—5	4	20—25	25
5—10	6	25—30	22
10—15	10	30—35	18
15—20	10	35—40	5

(Ans. 24; 17·5; 29·54)

10. From the table given below, find out the median and the quartiles :

Size	11—15	16—20	21—25	26—30	31—35	36—40
Freq.	7	10	13	26	35	40
Size	41—45	46—50				
Freq.	11	5				

(Ans. 33, 26·8, 37·9)

11. Find the Quartiles, 20th percentiles and the 8th decile of hts from the following table :

ht (in inches)	*No. of students*	*ht (in inches)*	*No. of students*
58	15	63	22
59	20	64	20
60	32	65	10
61	35	66	8
62	33		

(Ans. 60; 63; 60 and 63)

12. Find out the median and quartiles of data in Ex. 7.

(Ans. 12·8; 10·02; 16·4)

13. Find out the median, quartiles, 6th decile, 70th percentile and 3rd quartile for the data in Ex. 1-6.

(Ans. 124·75; 114·3; 136·47; 128·5; 133·53 and 21·85)

14. The following table gives the dist of farms according to their sizes in a given region. Calculate the median and the quartiles (size of the farm is rounded to the nearest acre) :

Farm size (acres)	*No. of farms*	*Farm size (acres)*	*No. of farms*
0—40	394	161—200	169
41—80	461	201—240	113
81—120	391	241 and over	148
121—160	334		

(Ans. 95·85; 49·91; 151·82)

15. Find mode from the follow
Wage (in Rs.) 20
No. of workers 8

16. Find mode of data in solve
17. Find mode of the data in E
18. Find mode of data in Ex. 1-
19. Find mode of data in Ex. 1-
20. Find mode of data given be
5 students get less than 3
12 students get less than 6
25 students get less than 9
30 students get less than 12
21. The consumption of petrol
planes to a hill station and
average would you conside
per gallon for up and down

22. Under what conditions wei
(*i*) equal to simple aver
(*ii*) greater than simple
(*iii*) less than simple ave
Illustrate your answer with
23. The following is the dist.
measure of central tendenc

Age-group	*No. of*
0—9	
10—19	
20—29	
30—39	

24. The table below shows the
2002.

Age	*No. (in mi*
under 25	2·2
25—29	4·0
30—34	5·0
35—44	10·4
45—54	9·4

Do you think that in this ca
mean ? Give reasons.
25. The daily expenditure of 1
Expenditure : 0—10
No. of families : 14
The median and mode for t
the missing frequencies.

absent	No. of students
n 30	487
n 35	493
n 40	497
n 45	500

(**Ans.** 13·51)

30	30—35	35—40
	58	49
55		

(**Ans.** 32·07)

rtiles :

lo. of students
25
22
18
5

(**Ans.** 24; 17·5; 29·54)

quartiles :

0	31—35	36—40
	35	40

(**Ans.** 33, 26·8, 37·9)

hts from the following table :

)	No. of students
	22
	20
	10
	8

(**Ans.** 60; 63; 60 and 63)

(**Ans.** 12·8; 10·02; 16·4)

e and 3rd quartile for the data

47; 128·5; 133·53 and 21·85)

their sizes in a given region.

rounded to the nearest acre) :

'acres)	No. of farms
00	169
40	113
l over	148

(**Ans.** 95·85; 49·91; 151·82)

15. Find mode from the following data :

Wage (in Rs.)	20	21	22	23	24	25	26	27	28
No. of workers	8	10	11	16	20	25	15	9	6

(**Ans.** 25)

16. Find mode of data in solved Ex. 1-4. (**Ans.** 17·78)

17. Find mode of the data in Ex. 7. (**Ans.** 12·48)

18. Find mode of data in Ex. 1-5. (**Ans.** 110·91)

19. Find mode of data in Ex. 1-6. (**Ans.** 123·41)

20. Find mode of data given below :

5 students get less than 3 marks
12 students get less than 6 marks
25 students get less than 9 marks
30 students get less than 12 marks (**Ans.** 7·29)

$Z = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$

21. The consumption of petrol by a motor was 'a gallon' for 20 k.m. while going up from planes to a hill station and 'a gallon' for 24 miles while coming down. What particular average would you consider appropriate for finding the average consumption in miles per gallon for up and down journey and why ?

$\left(\textbf{Ans. } 21\frac{9}{11} \text{ m.p.h. per gallon}\right)$

22. Under what conditions weighted average is

(*i*) equal to simple average.
(*ii*) greater than simple average.
(*iii*) less than simple average.

Illustrate your answer with the help of examples.

23. The following is the dist. of 136 individuals by 10-year age groups. Calculate that measure of central tendency which will appropriately describe the dist.

Age-group	*No. of persons*	*Age-group*	*No. of persons*
0—9	48	40—49	13
10—19	26	50—59	4
20—29	27	60—69	3
30—39	11	70 and over	4

(Median 17·2)

24. The table below shows the age dist of heads of families in country A during the year 2002.

Age	*No. (in millions)*	*Age*	*No. (in millions)*
under 25	2·22	55—64	6·63
25—29	4·05	65—74	4·16
30—34	5·08	75 and over	1·66
35—44	10·45		
45—54	9·47		

Do you think that in this case median is a better measure of central tendency than the mean ? Give reasons.

25. The daily expenditure of 100 families is given as under :

Expenditure :	0—10	10—20	20—30	30—40	40—50
No. of families :	14	?	27	?	15

The median and mode for the distribution are Rs. 25 and Rs. 29 respectively. Calculate the missing frequencies.

(**Ans.** 33, 11)

□□

2

Measures of Dispersion and Skewness

2.1. Introduction

In the preceding Chapter several measures used to describe the central tendency of a frequency distribution were discussed. These measures have their limitations and may conceal much pertinent factual information. It is also possible that these measures of central tendency may give results which are quite misleading. Thus, a measure of central tendency alone is not enough to give a correct picture of a distribution and for this some additional information is required. The following information is needed :

(1) The extent of scatteredness of items around central tendency. This is called *dispersion*.

(2) The direction of scatteredness. This is called *skewness*.

(3) The extent to which the distribution is more peaked or more flat-topped than the normal distribution. This is called *kurtosis*.

2.2. Measures of Dispersion

The object of measuring dispersion is to obtain a single summary figure which adequately exhibits the extent of the scatter of the variable values. Various measures of dispersion are :

(1) **Range, Interquartile range and Quartile deviation**

Range. It is difference between the greatest and least values of the variate.

Interquartile range. It is the difference between the upper and lower quartiles, *i.e.*, $Q_3 - Q_1$.

Quartile Deviation. *It is defined to be* $\frac{Q_3 - Q_1}{2}$, *where* Q_3 *and* Q_1 *are quartiles.*

Quartile Co-efficient of Dispersion. *It is defined to be*

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}.$$

Ex. 2-1. *Compute Quartile Deviation and the co-efficient of dispersion from the following data :*

Size	*Frequency*	*Size*	*Frequency*
4—8	6	24—28	12
8—12	10	28—32	10
12—16	18	32—36	6
16—20	30	36—40	2
20—24	15		

Sol.

Size C Inter
4—
8—
12—
16—
20—
24—
28—
32—
36—

Q_1 has $\frac{109}{4} = 27 \cdot 25$ it

∴ It lies in 12–16

∴

Q_3 has $\frac{3}{4}(109) = 81 \cdot 7$

∴ It lies in 24—28

∴

∴

and Co-efficient of dispersic

Sol.

Size Class Interval	*Frequency*	*C. Frequency*
4—8	6	6
8—12	10	16
12—16	18	34
16—20	30	64
20—24	15	79
24—28	12	91
28—32	10	101
32—36	6	107
36—40	2	109

Q_1 has $\frac{109}{4} = 27\cdot25$ items below it

∴ It lies in 12–16

$$\therefore \quad Q_1 = 12 + \frac{4}{18}(27\cdot25 - 16)$$

$$= 12 + \frac{4}{18}(11\cdot25)$$

$$= 12 + \frac{45}{18}$$

$$= 12 + 2\cdot5$$

$$= 14\cdot5.$$

Q_3 has $\frac{3}{4}(109) = 81\cdot75$ items below it.

∴ It lies in 24—28

$$\therefore \quad Q_3 = 24 + \frac{4}{12}(81\cdot75 - 79)$$

$$= 24 + \frac{1}{3}(2\cdot75)$$

$$= 24 + 0\cdot92 = 24\cdot92$$

$$\therefore \quad Q.D. = \frac{Q_3 - Q_1}{2} = \frac{24\cdot92 - 14\cdot5}{2}$$

$$= \frac{10\cdot42}{2} = 5\cdot21$$

and Co-efficient of dispersion

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{24\cdot92 - 14\cdot5}{24\cdot92 + 14\cdot5}$$

$$= \frac{10 \cdot 42}{39 \cdot 42} = 0 \cdot 26$$

(2) **Mean Deviation.** *It is defined by* M.D. $= \frac{1}{N} \sum f|x-a|$ *where 'a' is a point from which the deviations are to be taken.*

Co-efficient of Mean-Deviation. *It is defined to be*

$$\frac{\text{Mean-deviation about } 'a'}{a}$$

If nothing is mentioned usually mean deviation about median should be calculated.

Short Cut Method. *M.D. is calculated more easily by the formulae*

$$\text{M.D. about } 'a' = \frac{\sum f|x-b| + (a-b)\left(\sum_{x<a} f - \sum_{x>a} f\right)}{N}$$

$$\text{and M.D. about median } (M) = \frac{1}{N}\left\{\sum_{x>M} fx - \sum_{x<M} fx\right\}$$

Ex. 2-2. *Find the mean deviation for the following data :*

Height (in cms.)	*No. of students*	*Height (in cms.)*	*No. of students*
158	15	159	20
160	32	161	35
162	33	163	22
164	20	165	10
166	8		

Sol. It is not given about which the mean deviation is to be calculated. So mean deviation about median is to be calculated.

Calculation of Mean Deviation

Height x	*No. of Students Freq.* (f)	*c.f.*	$d = \lvert x-161 \rvert$	*fd*
158	15	15	3	45
159	20	35	2	40
160	32	67	1	32
161	35	102	0	0
162	33	135	1	33
163	22	157	2	44
164	20	177	3	60
165	10	187	4	40
166	8	195	5	40
	195			334

$$\text{Median} = \text{Value of } \left(\frac{195+1}{2}\right) \text{th item}$$

$$= \text{Value of 98th item}$$
$$= 161$$

26

$a|$ *where 'a' is a point from*

lian should be calculated.
: formulae

No. of students
20
35
22
10

alculated. So mean deviation

$x-161|$	fd
3	45
2	40
1	32
0	0
1	33
2	44
3	60
4	40
5	40
	334

em

$$\text{Mean Deviation} = \frac{334}{195}$$
$$= 1{\cdot}71.$$

Theorem 2.2-1. *Show that the mean deviation from the median is less than that measured from any other value.*

Sol. Let x be the variable.

Let $x_1, x_2, \ldots\ldots x_n$ be the values arranged in ascending order. Let M be the median. Then by def.,

$$\text{Mean deviation about } M = \frac{1}{n}\sum_{i=1}^{n}|x_i - M|$$
$$= \frac{1}{n}\sum_{x<M}|x-M| + \frac{1}{n}\sum_{x>M}|x-M|$$
$$= \frac{1}{n}\sum_{x<M}(M-x) + \frac{1}{n}\sum_{x>M}(x-M)$$
$$= \frac{1}{n}\sum_{x<M}(M-a+a-x) + \frac{1}{n}\sum_{x>M}(x-a+a-M)$$

where 'a' is any other point.

$$\therefore \text{ M.D. about } M = \frac{1}{n}\sum_{x<M}(M-a) + \frac{1}{n}\sum_{x<M}(a-x) + \frac{1}{n}\sum_{x>M}(x-a) + \frac{1}{n}\sum_{x>M}(a-M)$$
$$= \frac{1}{n}\sum_{x<M}(M-a) + \frac{1}{n}\sum_{x<M}(a-x) + \frac{1}{n}\sum_{x>M}(x-a) - \frac{1}{n}\sum_{x>M}(M-a)$$

As M is the median, the number of items for which $x < M$ is equal to the number of items for which $x > M$.

$$\therefore \qquad \sum_{x<M}(M-a) = \sum_{x>M}(M-a)$$

$$\therefore \qquad \text{M.D. about } M = \frac{1}{n}\sum_{x<M}(a-x) + \frac{1}{n}\sum_{x>M}(x-a)$$

(i) Let $a < M$

$$= \frac{1}{n}\sum_{x<a}(a-x) + \frac{1}{n}\sum_{a<x<M}(a-x) + \frac{1}{n}\sum_{x>a}(x-a) - \frac{1}{n}\sum_{M>x>a}(x-a)$$
$$= \left\{\frac{1}{n}\sum_{x<a}(a-x) + \frac{1}{n}\sum_{x>a}(x-a)\right\} - \frac{2}{n}\sum_{a<x<M}(x-a)$$
$$= \frac{1}{n}\sum_{i=1}^{n}|x_1 - a| - \frac{2}{n}\sum_{a<x<M}(x-a)$$

As in second term, $x > a$,

$$\frac{2}{n}\sum_{a<x<M}(x-a)$$

is non-negative.

$\therefore$ Mean deviation about median is less than that measured from any other value 'a'.

(*ii*) Let $a > M$.

Here M.D. about $M = \frac{1}{n}\sum_{x<M}(a-x)+\frac{1}{n}\sum_{x>M}(x-a)$

$$=\frac{1}{n}\sum_{M<x<a}(a-x)-\frac{1}{n}\sum_{M<x<a}(a-x)+\frac{1}{n}\sum_{x>a}(x-a)+\frac{1}{n}\sum_{a>x>M}(x-a)$$

$$=\frac{1}{n}\sum_{i=1}^{n}|x_i-a|-\frac{2}{n}\sum_{M<x<a}(a-x).$$

Since second term is non-negative, mean deviation about median is less than that measured from any other value 'a'.

Ex. 2-3. *Find the mean deviation about median from the following data :*

S.N.	*Marks*	*S.N.*	*Marks*	*S.N.*	*Marks*
1	17	7	41	13	11
2	32	8	32	14	15
3	35	9	11	15	35
4	33	10	18	16	23
5	15	11	20	17	38
6	21	12	22	18	12

(*i*) *by direct method.*

(*ii*) *by short cut method.*

Sol. Arranging Marks in ascending order :

S.N.	*Marks (x)*	$\lvert x-21\cdot5\rvert$	*S.N.*	*Marks (x)*	$\lvert x-21\cdot5\rvert$
1	11	10·5	10	22	0·5
2	11	10·5	11	23	1·5
3	12	9·5	12	32	10·5
4	15	6·5	13	32	10·5
5	15	6·5	14	33	11·5
6	17	4·5	15	35	13·5
7	18	3·5	16	35	13·5
8	20	1·5	17	38	16·5
9	21	0·5	18	41	19·5
	140			291	151·0

$$\text{Median} = \text{Value of } \frac{18+1}{2} = 9\cdot5\text{th item}$$

$$= \frac{\text{Value of 9th item} + \text{Value of 10th item}}{2}$$

ed from any other value 'a'.

$$+\frac{1}{n}\sum_{a>x>M}(x-a)$$

out median is less than that

following data :

S.N.	*Marks*
13	11
14	15
15	35
16	23
17	38
18	12

arks (x)	$\lvert x-21\cdot5\rvert$
22	0·5
23	1·5
32	10·5
32	10·5
33	11·5
35	13·5
35	13·5
38	16·5
41	19·5
291	151·0

item

alue of 10th item

$$=\frac{21+22}{2}=21\cdot5.$$

(*i*) By Direct method,
Mean deviation about median

$$=\frac{151}{18}=8\cdot4$$

(*ii*) By short cut method,
Mean deviation about median

$$=\frac{1}{N}\left\{\sum_{x>M}fx-\sum_{x<M}fx\right\}$$

Now $\sum_{x>M}fx=291$ and $\sum_{x<M}fx=140$

$\therefore$ M.D. about median $=\frac{151}{18}=8\cdot4.$

Ex. 2-4. *Show that the mean deviation about the mean $\bar{x}$ of the variate x can be written in the form.*

$$\frac{2}{N}\left[\bar{x}\sum_{x_i<\bar{x}}f_i-\sum_{x_i<\bar{x}}f_ix_i\right]$$

where f_i is the frequency of the value x_i.

Sol. Mean deviation about mean is given by

$$S=\frac{1}{N}\sum_i f_i\lvert x_i-\bar{x}\rvert$$

$$=\frac{1}{N}\sum_{x_i<\bar{x}}f_i(\bar{x}-x_i)+\frac{1}{N}\sum_{x_i>\bar{x}}f_i(x_i-\bar{x})$$

Now $\sum_i f_i(x_i-\bar{x})=\sum_i f_ix_i-\bar{x}\sum_i f_i=N\bar{x}-N\bar{x}=0$

$$\therefore \sum_{x_i<\bar{x}}f_i(x_i-\bar{x})+\sum_{x_i>\bar{x}}f_i(x_i-\bar{x})=0$$

or $$\sum_{x_i>\bar{x}}f_i(x_i-\bar{x})=\sum_{x_i<\bar{x}}f_i(\bar{x}-x_i)$$

$$\therefore S=\frac{2}{N}\sum_{x_i<\bar{x}}f_i(\bar{x}-x_i)=\frac{2}{N}\left[\bar{x}\sum_{x_i<\bar{x}}f_i-\sum_{x_i<\bar{x}}f_ix_i\right]$$

(3) Variance. It is defined by

$$\mu_2=\frac{1}{N}\Sigma f(x-\bar{x})^2$$

Standard Deviation. *It is the positive square root of the variance.*

Mean Square Deviation. *Mean square deviation about the pt 'a' is defined by*

$$\mu_2'(a) = \frac{1}{N} \Sigma f (x-a)^2.$$

Root Mean Square Deviation. *It is the positive square root of mean square deviation.*

Co-efficient of Variation. *It is defined to be*

$$100 \times \frac{(s.d.)}{mean}$$

Co-efficient of Dispersion. *It is defined by* $\dfrac{s.d.}{mean}$

For a given data s.d. is obtained by the formula

$$s.d. = h\sqrt{\frac{1}{N}\Sigma f.X^2 - \left(\frac{1}{N}\Sigma f X\right)^2}$$

where $$X = \frac{x-a}{h}.$$

Ex. 2-5. *Calculate the mean and s.d. of the following values of the world's annual gold output (in millions of pounds) for 20 different years :*

94	95	96	93	87	79	73	69	68	67
78	82	83	89	95	103	108	117	130	97

Also calculate the percentage of cases lying outside the mean at distances $\pm\sigma, \pm 2\sigma, \pm 3\sigma$ *where* σ *denotes the s.d.*

Sol. Arranging the data in ascending order :

Output (x) Arranged in order	$X = x - 90$	X^2	*Output (x) Arranged in order*	$X = x - 90$	X^2
67	–23	529	93	3	9
68	–22	484	94	4	16
69	–21	441	95	5	25
73	–17	289	95	5	25
78	–12	144	96	6	36
79	–11	121	97	7	49
82	–8	64	103	13	169
83	–7	49	108	18	324
87	–3	9	117	27	729
89	–1	1	130	40	1600
	–125	2131		128	2982

$$\therefore \quad \Sigma X = 128 - 125 = 3$$

$$\Sigma X^2 = 2982 + 2131 = 5113$$

$$\therefore \quad \text{A.M.} = 90 + \left(\frac{3}{20}\right) = 90 + 0{\cdot}15$$

$$= 90{\cdot}15 \text{ million pounds}$$

Now, me

$\therefore$ No. of cases outside t

$\therefore$ Percentage of cases o

No. of cases outside the

mean $\pm 2\sigma$

$\therefore$ Percentage of cases o

No. of cases outside the

i.e., $90{\cdot}15 \pm 47{\cdot}97$ or

$\therefore$ Percentage of cases o

Ex. 2-6. *The distribution factory is shown below. Comp the distribution :*

Max. Lo

9·3-

9·8-

10·3-

10·8-

11·3-

11·8-

12·3-

12·8-

$$S.D. = \sqrt{\frac{1}{n}\Sigma X^2 - \left(\frac{1}{n}\Sigma X\right)^2}$$

$$= \sqrt{\frac{1}{20}(5113) - \left(\frac{1}{20}\cdot 3\right)^2}$$

$$= \frac{1}{20}\sqrt{102260 - 9}$$

$$= \frac{1}{20}\sqrt{102251}$$

$$= \frac{319\cdot767}{20} = 15\cdot99 \text{ million pounds.}$$

Now, mean $\pm\sigma = 90\cdot15 \pm 15\cdot99$

$= 106\cdot14, 74\cdot16$

$\therefore$ No. of cases outside the range $74\cdot16$ to $106\cdot14 = 7$

$\therefore$ Percentage of cases outside the mean at distances $\pm\sigma$

$$= \frac{7}{20} \times 100 = 35\%.$$

No. of cases outside the range.

mean $\pm 2\sigma$ *i.e.*, $90\cdot15 \pm 31\cdot98$ or $58\cdot17$ to $122\cdot13 = 1$.

$\therefore$ Percentage of cases outside the mean at distances $\pm 2\sigma$

$$= \frac{1}{20} \times 100 = 5\%$$

No. of cases outside the range mean $\pm 3\sigma$

i.e., $90\cdot15 \pm 47\cdot97$ or $42\cdot18$ to $138\cdot12 = 0$

$\therefore$ Percentage of cases outside mean $\pm 3\sigma = 0\%$.

Ex. 2-6. *The distribution of maximum loads in tons supported by cables produced in a factory is shown below. Compute the standard deviation and the co-efficient of variation of the distribution :*

Max. Load (in tons)	*No. of cables*
9·3—9·7	2
9·8—10·2	5
10·3—10·7	12
10·8—11·2	17
11·3—11·7	14
11·8—12·2	6
12·3—12·7	3
12·8—13·2	1
	60

Sol.

Class intervals	*Freq* (f)	*Mid-points* (x)	$X = x - 11\cdot0$	$u = \dfrac{X}{0\cdot5}$	uf	$u^2 f$
9·3— 9·7	2	9·5	−1·5	−3	−6	18
9·8—10·2	5	10·0	−1·0	−2	−10	20
10·3—10·7	12	10·5	−0·5	−1	−12	12
10·8—11·2	17	11·0	0	0	0	0
11·3—11·7	14	11·5	0·5	1	14	14
11·8—12·2	6	12·0	1·0	2	12	24
12·3—12·7	3	12·5	1·5	3	9	27
12·8—13·2	1	13·0	2·0	4	4	16
	60				11	131

$$\text{S.D.} = (0.5)\sqrt{\frac{131}{60} - \left(\frac{11}{60}\right)^2}$$

$$= \frac{(0\cdot5)}{60}\sqrt{7860 - 121}$$

$$= \frac{1}{120}\sqrt{7739}$$

$$\log_{10}(\text{S.D.}) = \frac{1}{2}\log_{10}(7739) - \log_{10}(120)$$

$$= \frac{1}{2}(3\cdot8887) - 2\cdot0792$$

$$= 1\cdot94435 - 2\cdot0792$$

$$\simeq \bar{1}\cdot8652$$

$$\therefore \quad \text{S.D.} = \cdot07331 \simeq 0\cdot733 \text{ tons}$$

$$\text{A.M.} = 11\cdot0 + (0\cdot5)\left(\frac{11}{60}\right)$$

$$= 11 + \frac{11}{120} = \frac{1331}{120}$$

$$= 11\cdot092 \text{ tons.}$$

∴ Co-efficient of variation

$$= \left(\frac{\text{S.D.}}{\text{A.M.}}\right)(100)$$

$$= \left(\frac{0\cdot733}{11\cdot092}\right)(100)$$

$$= \frac{73300}{11092} = 6\cdot61\%$$

Ex. 2-7. (*a*) *Find out the co-ef*

Va

Mea

(*b*) *If in a series which is not h*

approximate value of its s.d. (use l

Sol. (*a*) S.D. = $\sqrt{148\cdot6} = 12\cdot1$

Co-efficient of variation

(*b*) M.D

But M.D

∴ S.D

Ex. 2-8. *Calculate s.d. of data gi*
frequency of less than type) :

Marks	*St*
80—84	
75—79	
70—74	
65—69	
60—64	
55—59	

Sol.

Mid points (x)	*Freq.* (f)
27	1
32	0
37	3
42	6
47	6
52	6
57	7
62	4

$u = \dfrac{X}{0 \cdot 5}$	uf	$u^2 f$
−3	−6	18
−2	−10	20
−1	−12	12
0	0	0
1	14	14
2	12	24
3	9	27
4	4	16
	11	131

(120)

Ex. 2-7. (*a*) *Find out the co-efficient of variation if*

$$Var = 148{\cdot}6$$
$$Mean = 40$$

(*b*) *If in a series which is not highly skewed the mean deviation 7·8, what would be the approximate value of its s.d.* $\left(\text{use M.D.} = \dfrac{4}{5}\text{s.d.}\right)$

Sol. (*a*) S.D. $= \sqrt{148 \cdot 6} = 12 \cdot 19$

Co-efficient of variation

$$= \left(\frac{\text{S.D.}}{\text{A.M.}}\right)(100).$$

$$= \left(\frac{12 \cdot 19}{40}\right)(100) = \frac{1219}{40}$$

$$= 30 \cdot 5\%$$

(*b*) $\qquad$ M.D. $= 7{\cdot}8$

But $\qquad$ M.D. $= \dfrac{4}{5}$(S.D.)

$\therefore \qquad$ S.D. $= \dfrac{5}{4}$ (M.D.) $= \dfrac{5}{4}(7{\cdot}8)$

$$= \frac{39}{4} = 9 \cdot 75.$$

Ex. 2-8. *Calculate s.d. of data given below by the method of summation (using cumulative frequency of less than type) :*

Marks	*Students*	*Marks*	*Students*
80—84	1	50—54	6
75—79	1	45—49	6
70—74	1	40—44	6
65—69	4	35—39	3
60—64	4	30—34	0
55—59	7	25—29	1
			40

Sol.

Mid points (*x*)	*Freq.* (*f*)	*First cumulation* (*c. Freq.*)	*Second cumulation* (*c. Freq. of c. Freq.*)
27	1	1	1
32	0	1	2
37	3	4	6
42	6	10	16
47	6	16	32
52	6	22	54
57	7	29	83
62	4	33	116

(*Contd.*)

67	4	37	153
72	1	38	191
77	1	39	230
82	1	40	270
	40	270	1154

Standard deviation by the method of summation is given by

$$SD = h\sqrt{2F_2 - F_1 - F_1^2}$$

where h = common class-interval

F_1 = The Sum of Cumulative Frequencies (less than type) divided by the number of items.

F_2 = The sum of Cumulative Frequencies (less than type) of the Cumulative Frequencies (less than type) divided by the number of items.

Here $h = 5$

$$F_1 = \frac{270}{40}$$

$$F_2 = \frac{1154}{40}$$

$$\therefore \quad S.D. = 5\sqrt{\frac{2308}{40} - \frac{270}{40} - \left(\frac{270}{40}\right)^2}$$

$$= \frac{1}{8}\sqrt{92320 - 10800 - 72900}$$

$$= 11{\cdot}61.$$

Ex. 2-9. *Calculate the s.d. of data given below by the method of summation (using more than type cumulative frequency).*

Sol.

Mid-points (x)	*Freq. (f)*	*First Cumulation (c. Freq.)*	*Second Cumulation (c. Freq. of c. Freq.)*
75	12	230	1045
85	18	218	815
95	35	200	597
105	42	165	397
115	50	123	232
125	45	73	109
135	20	28	36
145	8	8	8
	230	1045	3239

S.D. by the method of summation is given by

$$S.D. = h\sqrt{2F_2 - F_1^2 - F_1}$$

where h = common class-interval

F_1 = The sum of cumulative F
items.

F_2 = The sum of cumulative Fre
(more than type) divided by the nu

Here

F_1

F_2

$\therefore$ S.D

Ex. 2-10. *The following table*
companies A and B. Find out which

Share A : 318 322

Share B : 2542 2542

Sol. Arranging the values in as

Share A	
x	$d_1 = x - 318$
308	−10
312	− 6
315	− 3
318	0
319	1
322	4
324	6
325	7
	− 1

For Share A,

A.M.

S.D.

	153
	191
	230
	270
	1154

iven by

erval
ılative Frequencies (less than type)
mber of items.
pe) of the Cumulative Frequencies

$\left(\frac{270}{40}\right)^2$

) – 72900

'he method of summation (using

	Second Cumulation (c. Freq. of c. Freq.)
	1045
	815
	597
	397
	232
	109
	36
	8
	3239

F_1 = The sum of cumulative Frequencies (more than type) divided by the number of items.

F_2 = The sum of cumulative Frequencies (more than type) of the cumulative frequencies (more than type) divided by the number of items.

Here $\qquad h = 10$

$$F_1 = \frac{1045}{230}$$

$$F_2 = \frac{3239}{230}$$

$$\therefore \qquad S.D. = 10\sqrt{\frac{2(3239)}{230} - \frac{1045}{230} - \left(\frac{1045}{230}\right)^2}$$

$$= \frac{1}{23}\sqrt{(6478)(230) - (1045)(230) - (1045)^2}$$

$$= \frac{1}{23}\sqrt{1489940 - 240350 - 1092025}$$

$$= \frac{1}{23}\sqrt{157565} = 17 \cdot 258.$$

Ex. 2-10. *The following table gives the fluctuations in the prices of shares of two companies A and B. Find out which of them shows greater variability ?*

Share A :	318	322	325	312	324	315	308	319
Share B :	2542	2542	2534	2532	2545	2530	2566	2550

Sol. Arranging the values in ascending order :

Share A			*Share B*		
x	$d_1 = x - 318$	d_1^2	y	$d_2 = y - 2542$	d_2^2
308	−10	100	2530	−12	144
312	− 6	36	2532	−10	100
315	− 3	9	2534	− 8	64
318	0	0	2542	0	0
319	1	1	2542	0	0
322	4	16	2545	3	9
324	6	36	2550	8	64
325	7	49	2566	24	576
	− 1	247		5	957

For Share A,

$$A.M. = 318 - \frac{1}{8} = 318 - 0 \cdot 125$$

$$= 317 \cdot 875$$

$$S.D. = \sqrt{\frac{247}{8} - \left(-\frac{1}{8}\right)^2}$$

$$= \frac{1}{8}\sqrt{1976-1} = \frac{1}{8}\sqrt{1975}$$

$$= \frac{1}{8}(44 \cdot 441) = 5 \cdot 555$$

$$\text{Co-efficient of variation} = \left(\frac{S.D.}{A.M.}\right)(100)$$

$$= \left(\frac{5 \cdot 555}{317 \cdot 875}\right)(100)$$

$$= 17{\cdot}5\%.$$

For Share B,

$$A.M. = 2542 + \frac{5}{8} = 2542 \cdot 625$$

$$S.D. = \sqrt{\frac{957}{8} - \left(\frac{5}{8}\right)^2}$$

$$= \frac{1}{8}\sqrt{7656-25} = \frac{1}{8}\sqrt{7631}$$

$$= \frac{1}{8}(87 \cdot 36) \simeq 10 \cdot 92$$

$$\text{Co-efficient of variation} = \left(\frac{10 \cdot 92}{2542 \cdot 625}\right)(100)$$

$$= \frac{1092000}{2542625} \doteq 0 \cdot 43\%.$$

Since co-efficient of variation for share A is greater than that for share B, share A shows greater variability.

Ex. 2-11. *On a final examination in statistics, the mean marks of a group of 150 students were 78 and the s.d. was 8·0. In Economics, however, the mean marks of the group were 73 and the s.d. was 7·6. In what subject was there greater variability ?*

Sol. Co-efficient of variation for statistics paper

$$= \left(\frac{8 \cdot 0}{78}\right)(100) = \frac{800}{78}$$

$$= 10{\cdot}3\%.$$

Co-efficient of variation for Economics paper

$$= \left(\frac{7 \cdot 6}{73}\right)(100) = \frac{760}{73}$$

$$= 10{\cdot}4\%$$

$\therefore$ In Economics there was greater variability.

Ex. 2-12. Show that if the vari

tional to the binomial co-efficients

the dist is $\frac{n}{2}$, the mean square dev

Sol. $A.M$

μ_2' ((

$\therefore$

Ex. 2-13. *Find the mean devi*

a+2nd, and prove that the latter i

Sol. A.M. is given by

$\sqrt{1975}$

555

·625

$\sqrt{7631}$

2

)

6.

n that for share B, share A shows

marks of a group of 150 students
nean marks of the group were 73
riability ?

Ex. 2-12. Show that if the variable takes the values 0, 1, 2,...n with frequencies proportional to the binomial co-efficients 1, ${}^nc_1,\ {}^nc_2, \ldots\ldots\ldots\ {}^nc_n$ respectively, then the mean of the dist is $\frac{n}{2}$, the mean square deviation about $x=0$ is $\frac{n(n+1)}{4}$ and the variance is $\frac{n}{4}$.

Sol.

$$A.M. = \frac{0.1 + 1.{}^nc_1 + 2.{}^nc_2 + 3.{}^nc_3 + \ldots\ldots + n.{}^nc_n}{1 + {}^nc_1 + {}^nc_2 + \ldots\ldots + {}^nc_n}$$

$$= \frac{n\left\{1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \ldots + 1\right\}}{(1+1)^n}$$

$$= n\left\{\frac{1 + {}^{n-1}c_1 + {}^{n-1}c_2 + \ldots + {}^{n-1}c_{n-1}}{2^n}\right\}$$

$$= n.\frac{2^{n-1}}{2^n} = \frac{n}{2}$$

$$\mu_2'(0) = \frac{0^2.1 + 1^2.{}^nc_1 + \ldots + n^2.{}^nc_n}{2^n}$$

$$= \frac{1}{2^n}\sum_{x=0}^{n} x^2.{}^nc_x = \frac{1}{2^n}\sum_{x=0}^{n}\{x(x-1)+x\}.{}^nc_x$$

$$= \frac{1}{2^n}\sum_{x=0}^{n} x(x-1).{}^nc_x + \frac{1}{2^n}\sum_{x=0}^{n} x.{}^nc_x$$

$$= \frac{1}{2^n}\{2.1.{}^nc_2 + 3.2.{}^nc_3 + \ldots + n(n-1){}^nc_n\} + \frac{n}{2}$$

$$= \frac{1}{2^n}n(n-1)(1+1)^{n-2} + \frac{n}{2} = \frac{n(n-1)}{4} + \frac{n}{2}$$

$$= \frac{n(n+1)}{4}$$

$$\therefore \quad \mu_2 = \frac{n(n+1)}{4} - \frac{n^2}{4} = \frac{n}{4}.$$

Ex. 2-13. *Find the mean deviation from the mean and the s.d. of the A.P. a, a+d,........ a+2nd, and prove that the latter is greater than the former.*

Sol. A.M. is given by

$$\bar{x} = \frac{a + (a+d) + \ldots + (a+2nd)}{2n+1} = a + d\left\{\frac{1+2+\ldots\ldots+2n}{2n+1}\right\}$$

$$= a + d\frac{2n(2n+1)}{2(2n+1)} = a + nd$$

$\therefore$ Mean deviation from the mean

$$= \frac{1}{2n+1}[\{|-nd|+|-(n-1)d|+\ldots.+|d|\}+\{|d|+\ldots.+|nd|\}]$$

$$= \frac{2d}{2n+1}\{1+2+\ldots\ldots+n\} = \frac{n(n+1)d}{(2n+1)}$$

$$S.D. = \sqrt{\frac{1}{2n+1}[\{(-nd)^2+(-(n-1)d)^2+\ldots(-d)^2\}+\{d^2+\ldots+n^2d^2\}]}$$

$$= \sqrt{\frac{2d^2}{2n+1}\{1^2+2^2+\ldots+n^2\}} = d\sqrt{\frac{2n(n+1)(2n+1)}{6(2n+1)}}$$

$$= d\sqrt{\frac{n(n+1)}{3}}.$$

Consider

$$(\text{Mean deviation})^2 - \text{Variance} = d^2\left\{\frac{n^2(n+1)^2}{(2n+1)^2} - \frac{n(n+1)}{3}\right\}$$

$$= \frac{n(n+1)d^2}{3(2n+1)^2}\{3n(n+1)-(4n^2+4n+1)\}$$

$$= -\frac{n(n+1)d^2}{3(2n+1)^2}(n^2+n+1) < 0$$

$\therefore$ Mean deviation $< S.D.$

Ex. 2-14. *If r be the range and $S = \left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}}$ be the s.d. of a set of observations $x_1, x_2, \ldots\ldots\ldots x_n$ then show that*

$$S \le r\left(\frac{n}{n-1}\right)^{\frac{1}{2}}$$

Sol. Let $x_r = \max.(x_1, x_2 \ldots x_n)$

and $x_k = \min.(x_1, x_2 \ldots\ldots. x_n)$

Then $r = x_r - x_k$

Now $\bar{x} = \frac{x_1+x_2+\ldots.+x_n}{n} \ge \frac{x_k+x_k+\ldots.+x_k}{n} = x_k$

$$\therefore \quad (x_i-\bar{x})^2 \le (x_i-x_k)^2$$

$$\therefore \; S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \le \frac{1}{n-1}\sum_{i=1}^{n}(x_i-x_k)^2$$

$\le \frac{}{n}$

$= \frac{nr^2}{n-1}$

$\therefore \quad S \le r\left\{\frac{n}{n-1}\right\}$

Theorem 2.2-2. *Show that the ... measured from the mean.*

Sol. Let 'a' be an arbitrary poi...

$\mu_2'(a) = \frac{1}{N}\sum_{i=1}^{n} f$

where $\bar{x}$ is the A.M.

$= \frac{1}{N}\sum_{i=1}^{n} f$

$= \frac{1}{N}\sum_{i=1}^{n} f_i($

$= \mu_2 + (\bar{x} -$

$\therefore \quad \mu_2'(a) - \mu_2 = (\bar{x}-a)^2$

or $\quad \mu_2'(a) \ge \mu_2$

$\therefore \quad \sqrt{\mu_2'(a)} \ge \sqrt{\mu_2}$

$\therefore$ The root mean square dev... mean.

Ex. 2-15. *In a series of measur... of magnitude x_2 and so on. If $\bar{x}$ is*

$S.D. = \sqrt{\frac{\Sigma m_r(k-x_r)^2}{\Sigma m_r} - \delta^2}$

where $\bar{x} = k + \delta$ and k is any const...

Sol. μ

$\therefore \quad S.D$

$+|d|\} + \{|d|+\dots+|nd|\}]$

$1)d$

$1)$

$^2+\dots(-d)^2\}+\{d^2+\dots+n^2d^2\}]$

$\sqrt{\frac{2n(n+1)(2n+1)}{6(2n+1)}}$

$1)\Big\}$

$n+1)\}$

$\bar{x})^2\Big\}^{\frac{1}{2}}$ *be the s.d. of a set of*

$$\leq \frac{1}{n-1}\left\{\sum_{i=1}^{n}(x_r - x_k)^2\right\}.$$

$$= \frac{nr^2}{n-1} \qquad (\because\ x_r - x_k = r)$$

$$\therefore \qquad S \leq r\left\{\frac{n}{n-1}\right\}^{\frac{1}{2}}$$

Theorem 2.2-2. *Show that the root mean square deviation is least when deviations are measured from the mean.*

Sol. Let 'a' be an arbitrary point. Then

$$\mu_2'(a) = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x} + \bar{x} - a)^2$$

where $\bar{x}$ is the A.M.

$$= \frac{1}{N}\sum_{i=1}^{n} f_i\left\{(x_i - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_i - \bar{x})(\bar{x} - a)\right\}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2 + (\bar{x} - a)^2 . \frac{1}{N}\sum_{i=1}^{n} f_i \ + 2(\bar{x} - a)\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})$$

$$= \mu_2 + (\bar{x} - a)^2$$

$$\therefore \qquad \mu_2'(a) - \mu_2 = (\bar{x} - a)^2 \geq 0$$

or $$\mu_2'(a) \geq \mu_2$$

$$\therefore \qquad \sqrt{\mu_2'(a)} \geq \sqrt{\mu_2}$$

$\therefore$ The root mean square deviation is least when deviations are measured from the mean.

Ex. 2-15. *In a series of measurements we obtain m_1 values of magnitude x_1, m_2 values of magnitude x_2 and so on. If $\bar{x}$ is the mean value of all the measurements, prove that*

$$S.D. = \sqrt{\frac{\Sigma m_r(k - x_r)^2}{\Sigma m_r} - \delta^2}$$

where $\bar{x} = k + \delta$ and k is any constant.

Sol. $$\mu_2 = \mu_2'(k) - (\bar{x} - k)^2 = \mu_2'(k) - \delta^2$$

$$\therefore \qquad S.D. = \sqrt{\frac{\Sigma m_r(k - x_r)^2}{\Sigma m_r} - \delta^2}$$

Theorem 2.2-3. *Show that S.D. is independent of origin but not of scale.*

Sol. The transformation corresponding to change of origin and scale is

$$U = \frac{x-a}{h}$$

where 'a' corresponds to change of origin and h to change in scale.

$$\therefore \qquad x = a + Uh$$

Then
$$\bar{x} = \frac{1}{N}\sum_{i-1}^{n} f_i(x_i) = \frac{1}{N}\sum_{i=1}^{n} f_i(a+U_i h)$$

$$= a + h\frac{1}{N}\sum_{i=1}^{n} f_i U_i = a + h\bar{U}$$

$$\therefore \qquad \mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\{x_i - \bar{x}\}^2 = h^2.\frac{1}{N}\sum_{i=1}^{n} f_i(U_i - \bar{U})^2$$

$$= h^2.\mu_2 \text{ for } U$$

$$\therefore \qquad \mu_2 \text{ for } x = h^2.\mu_2 \text{ for } U$$

$\therefore$ Variance and hence s.d. is independent of origin but not of scale.

Ex. 2-16. *From a sample of n observations, the A.M. and variance are calculated. It is then found that one of the values x_1 is in error and should be replaced by x_1'. Show that the adjustment to the variance to correct this error is*

$$\frac{1}{n}(x_1' - x_1)\left(x_1' + x_1 - \frac{x_1' - x_1 + 2T}{n}\right)$$

where T is the total of original observations.

Sol. Let $\bar{x}$ and o^2 be the calculated values of A.M. and variance.

Then
$$\Sigma x_i = n\bar{x}$$

and
$$\Sigma(x_i - \bar{x})^2 = n\sigma^2$$

i.e.,
$$\Sigma x_i^2 - n\bar{x}^2 = n\sigma^2$$

$$\therefore \qquad \Sigma x_i^2 = n\sigma^2 + n\bar{x}^2$$

Now corrected value of $\Sigma x_i = \left(\Sigma x_i - x_1 + x_1'\right)$

Corrected value of $\Sigma x_i^2 = \left(\Sigma x_i^2 - x_1^2 + x_1'^2\right)$

$\therefore$ Corrected value of A.M. $= \dfrac{1}{n}\{\Sigma x_i - x_1 + x_1'\}$

$$= \left(\bar{x} + \frac{x_1' - x_1}{n}\right)$$

and corrected value of $\frac{1}{n}\Sigma$

$\therefore$ Corrected value of varianc

$\therefore$ Adjustment to the varianc

Now

$\therefore$ Adjustment to the varianc

Ex. 2-17. *For a frequency di*
in intervals 0—5, 5—10, ...etc.)
discovered that the score 43 wa.
corrected mean and s.d. corresp.

Sol. Since the score 43 was r
45 and 50—55, in the calculation
of the actual value 42·5.

Now if x be the variate,

$\therefore$ Corrected value of

$\therefore$ Corrected value of mean

Also

$\therefore$ Va

$\therefore$ $\Sigma(x -$

igin but not of scale.
origin and scale is

ge in scale.

$= a + h\overline{U}$

$= h^2 . \frac{1}{N} \sum_{i=1}^{n} f_i (U_i - \overline{U})^2$

ut not of scale.
and variance are calculated. It is
be replaced by x_1' *. Show that the*

$\cdot 2T \big)$

nd variance.

and corrected value of $\quad \frac{1}{n} \Sigma x_i^2 = \frac{1}{n} \{\Sigma x_i^2 - x_1^2 + x_1'^2\}$

$$= \sigma^2 + \bar{x}^2 + \frac{x_1'^2 - x_1^2}{n}$$

$\therefore$ Corrected value of variance

$$= \sigma^2 + \bar{x}^2 + \frac{x_1'^2 - x_1^2}{n} - \left(\bar{x} + \frac{x_1' - x_1}{n} \right)^2$$

$\therefore$ Adjustment to the variance to correct the error

$$= \text{corrected value of variance } - \sigma^2$$

$$= \bar{x}^2 + \frac{x_1'^2 - x_1^2}{n} - \left(\bar{x} + \frac{x_1' - x_1}{n} \right)^2$$

$$= \frac{1}{n}(x_1' - x_1) \left\{ x_1' + x_1 - 2\bar{x} - \frac{x_1' - x_1}{n} \right\}$$

Now $\quad \bar{x} = \frac{1}{n} \Sigma x_i = \frac{T}{n} \qquad (\because T = \Sigma x_i)$

$\therefore$ Adjustment to the variance

$$= \frac{1}{n}(x_1' - x_1) \left\{ x_1' + x_1 - \frac{x_1' - x_1 + 2T}{n} \right\}$$

Ex. 2-17. *For a frequency distribution of marks in History of 200 candidates (grouped in intervals 0—5, 5—10, ...etc.) the mean and s.d. were found to be 40 and 15. Later it was discovered that the score 43 was misread as 53 in obtaining the frequency dist. Find the corrected mean and s.d. corresponding to the corrected frequency dist.*

Sol. Since the score 43 was misread as 53 and the scores 43 and 53 lie in intervals 40—45 and 50—55, in the calculation of mean and s.d. variate value was taken to be 52·5 instead of the actual value 42·5.

Now if x be the variate,

$$\Sigma x = (40)(200) = 8000$$

$\therefore$ Corrected value of

$$\Sigma x = 8000 - 52 \cdot 5 + 42 \cdot 5 = 7990$$

$\therefore$ Corrected value of mean

$$= \frac{7990}{200} = 39 \cdot 95$$

Also $\quad s.d. = 15$

$\therefore \quad \text{Var}(x) = 225$

$\therefore \quad \Sigma(x - \bar{x})^2 = (225)(200) = 45000$

or $$\Sigma x^2 - N\bar{x}^2 = 45000$$

$\therefore$ $$\Sigma x^2 = 45000 + (200)(1600) = 365000$$

$\therefore$ Corrected value of Σx^2

$$= 365000 - (52\cdot 5)^2 + (42\cdot 5)^2 = 364050$$

$\therefore$ Corrected value of $\Sigma(x-\bar{x})^2 =$ (Corrected value of $\Sigma x^2) - N$ (corrected mean)^2

$$= 364050 - 200(39\cdot 95)^2 = 44849\cdot 5$$

$\therefore$ Corrected variance

$$= \frac{44849\cdot 5}{200} = 224\cdot 2475$$

$\therefore$ Corrected s.d. $= \sqrt{224\cdot 2475} = 14\cdot 97.$

Ex. 2-18. *The mean of 5 observations is 4·4 and the variance is 8·24. If three of the five observations are 1, 2 and 6, find the other two.*

Sol. Let x_1 and x_2 be other observations. Then

$$(4\cdot 4)(5) = (1+2+6) + (x_1 + x_2)$$

or $$x_1 + x_2 = 22 - 9 = 13 \qquad \ldots(1)$$

and $$(8\cdot 24)(5) = (1-4\cdot 4)^2 + (2-4\cdot 4)^2 + (6-4\cdot 4)^2 + (x_1 - 4\cdot 4)^2 + (x_2 - 4\cdot 4)^2$$

$\therefore$ $$x_1^2 + x_2^2 = 97\cdot 0$$

Now $2(x_1^2 + x_2^2) = (x_1 + x_2)^2 + (x_1 - x_2)^2$

$\therefore$ $$x_1 - x_2 = 5 \quad \text{(taking positive sign)} \qquad \ldots(2)$$

From (1) and (2)

$$x_1 = 9, \quad x_2 = 4.$$

Ex. 2-19. *If the mean and s.d. of a variate x are m and σ respectively, obtain the mean and s.d. of* $\frac{ax+b}{c}$ *where a, b and c are constants.*

Sol. Let $U = \frac{ax+b}{c}$

Let $\bar{U}$ and σ_U be the mean and s.d. of U.

Then $$\bar{U} = \frac{1}{N}\sum f\left(\frac{ax+b}{c}\right) = \frac{1}{c}\left\{a\frac{1}{N}\sum fx + b\frac{1}{N}\sum f\right\}$$

$$= \frac{a\bar{x}+b}{c} = \frac{am+b}{c}$$

and $$\sigma_U^2 = \frac{1}{N}\sum f(U-\bar{U})^2 = \frac{a^2}{c^2}\frac{1}{N}\sum f(x-\bar{x})^2 = \frac{a^2}{c^2}\sigma^2$$

$$\therefore \sigma_U = \left|\frac{a}{c}\right|\sigma.$$

$)) = 365000$

$(42 \cdot 5)^2 = 364050$

$\Sigma x^2) - N$ (corrected mean)^2

$5)^2 = 44849 \cdot 5$

75

ance is 8·24. If three of the five

...(1)

$^2 + (x_1 - 4 \cdot 4)^2 + (x_2 - 4 \cdot 4)^2$

...(2)

respectively, obtain the mean

σ^2

Theorem 2.2-4. *Show that the s.d. is not less than the mean deviation from the mean.*

Sol. Let $\bar{x}$ be the A.M. Then it is to be proved that s.d.$\nless$ mean deviation from the mean.

i.e., S.D. $\geq$ Mean deviation from the mean

$$i.e., \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2} \geq \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - \bar{x}|$$

$$i.e., \quad N\sum_{i=1}^{n} f_i y_i^2 \geq \left(\sum_{i=1}^{n} f_i y_i\right)^2$$

where $\qquad y_i = |x_i - \bar{x}|$

$$i.e., (f_1 + f_2 + ... + f_n)(f_1 y_1^2 + ... + f_n y_n^2) \geq (f_1 y_1 + f_2 y_2 + ... + f_n y_n)^2$$

$$i.e., \qquad f_1 f_2 (y_1 - y_2)^2 + ... \geq 0$$

which is true.

Ex. 2-20. *Show that if the deviations are small compared with the mean so that* $\left(\frac{x}{M}\right)^3$ *and higher powers of* $\left(\frac{x}{M}\right)$ *may be neglected,*

(*i*) $G = M\left(1 - \frac{\sigma^2}{2M^2}\right)$ *where 'G' is the G.M. and 'M' the A.M. and 'σ' the s.a.*

(*ii*) $H = M\left(1 - \frac{\sigma^2}{M^2}\right)$ *where H is the H.M.*

(*iii*) $H + M = 2G$.

(*iv*) $M^2 - G^2 = \sigma^2$

(*v*) $MH = G^2$.

(*vi*) *mean* $(\sqrt{x}) = \sqrt{M}\left(1 - \frac{\sigma^2}{8M^2}\right)$.

Sol. (*i*) By def.

$$\log G = \frac{1}{N}\sum_{i=1}^{n} f_i \log x_i$$

Let $X_i = x_i - M$ so that $x_i = X_i + M$

$$\therefore \qquad \log G = \frac{1}{N}\sum_{i=1}^{n} f_i \log(X_i + M)$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \left\{\log M + \log\left(1+\frac{X_i}{M}\right)\right\}$$

$$= \log M + \frac{1}{N}\sum_{i=1}^{n} f_i \log\left(1+\frac{X_i}{M}\right)$$

Applying expansion of $\log\left(1+\frac{X_i}{M}\right)$ and neglecting $\left(\frac{X_i}{M}\right)^3$ and higher powers

$$\log G = \log M + \frac{1}{N}\sum_{i=1}^{n} f_i \left\{\frac{X_i}{M} - \frac{1}{2}\frac{X_i^2}{M^2}\right\}$$

$$= \log M + \frac{1}{M}\cdot\frac{1}{N}\sum_{i=1}^{n} f_i X_i - \frac{1}{2M^2}\cdot\frac{1}{N}\sum_{i=1}^{n} f_i X_i^2$$

$$\therefore \log\frac{G}{M} = -\frac{1}{2M^2}\sigma^2 \qquad \left(\because \sum_{i=1}^{n} f_i X_i = 0\right)$$

$$\therefore G = Me^{-\frac{1}{2M^2}\sigma^2} = M\left(1-\frac{\sigma^2}{2M^2}\right)$$

(Applying the expansion of $e^{-\frac{\sigma^2}{2M^2}}$ and neglecting higher powers)

(*ii*) By def. $\frac{1}{H} = \frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \frac{1}{X_i+M} = \frac{1}{M}\cdot\frac{1}{N}\sum_{i=1}^{n} f_i\left(1+\frac{X_i}{M}\right)^{-1}$$

$$= \frac{1}{M}\frac{1}{N}\sum_{i=1}^{n} f_i\left\{1-\frac{X_i}{M}+\frac{X_i^2}{M^2}\right\} = \frac{1}{M}\left(1+\frac{\sigma^2}{M^2}\right)$$

$$\therefore \qquad H = M\left\{1+\frac{\sigma^2}{M^2}\right\}^{-1} = M\left(1-\frac{\sigma^2}{M^2}\right)$$

(*iii*) From (*ii*) $H + M = M\left(2-\frac{\sigma^2}{M^2}\right) = 2M\left(1-\frac{\sigma^2}{2M^2}\right)$

$$= 2G$$

(*iv*) From (*i*) $G^2 = M^2\left(1-\frac{\sigma^2}{2M^2}\right)^2 = M^2\left(1-\frac{\sigma^2}{M^2}\right)$ (neglecting higher powers)

$\therefore\ M^2 - G^2 = \sigma^2$

(*v*) From (*ii*) MH $= M^2 - \sigma$

(*vi*) mean

Ex. 2-21. *Show that, if the*

$\left(\frac{x}{M}\right)^3$ *and higher powers may*

where V is the co-efficient of va

Sol. From last Ex.

$\therefore$ 2(M -

$\therefore$ V= co-efficient of vari

2.3. Combining number

$m_1, m_2 \ldots m_k$, *sizes* $n_1, n_2, \ldots\ldots$

mean and s.d. of the new distri

and

og $\left(1+\frac{X_i}{M}\right)\Big\}$

g $\left(1+\frac{X_i}{M}\right)$

$\left.\frac{X_i}{M}\right)^3$ and higher powers

$\left(\because \sum_{i=1}^{n} f_i X_i = 0\right)$

er powers)

$\cdot \frac{1}{N}\sum_{i=1}^{n} f_i\left(1+\frac{X_i}{M}\right)^{-1}$

$+\left.\frac{X_i^2}{M^2}\right\} = \frac{1}{M}\left(1+\frac{\sigma^2}{M^2}\right)$

$\left.1-\frac{\sigma^2}{M^2}\right)$

glecting higher powers)

$$= M^2 - \sigma^2$$

$\therefore\ M^2 - G^2 = \sigma^2$

(*v*) From (*ii*) MH $= M^2 - \sigma^2 = G^2$ [from (*iv*)]

(*vi*) mean $$(\sqrt{x}) = \frac{1}{N}\sum_{i=1}^{n} f_i \sqrt{x_i}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \{X_i + M\}^{\frac{1}{2}} = \sqrt{M}\,\frac{1}{N}\sum_{i=1}^{n} f_i\left(1+\frac{X_i}{M}\right)^{\frac{1}{2}}$$

$$= \sqrt{M}\cdot\frac{1}{N}\sum_{i=1}^{n} f_i\left\{1+\frac{1}{2}\frac{X_i}{M}-\frac{1}{8}\frac{X_i^2}{M^2}\right\}$$

$$= \sqrt{M}\cdot\left\{1-\frac{1}{8M^2}\sigma^2\right\}.$$

Ex. 2-21. *Show that, if the deviations are small compared with the mean M so that* $\left(\frac{x}{M}\right)^3$ *and higher powers may be neglected.*

$$V = \sqrt{\frac{2(M-G)}{M}}$$

where V is the co-efficient of variation.

Sol. From last Ex.

$$G = M\left(1-\frac{\sigma^2}{2M^2}\right)$$

$$\therefore \qquad 2(M-G) = \frac{\sigma^2}{M}$$

$$\therefore\ V = \text{co-efficient of variation} = \frac{\text{s.d.}}{\textit{mean}} = \frac{\sigma}{M}$$

$$= \sqrt{\frac{2(M-G)}{M}}$$

2.3. Combining number of distributions. *If k-distributions with respective means* $m_1, m_2 \ldots m_k$, *sizes* $n_1, n_2, \ldots\ldots n_k$ *and s.d.s* $\sigma_1, \sigma_2, \ldots\ldots \sigma_k$ *be combined together, the mean and s.d. of the new distribution are given by*

$$m = \frac{n_1 m_1 + n_2 m_2 + \ldots + n_k m_k}{n_1 + n_2 + \ldots + n_k}$$

and $$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2 + \ldots + n_k m_k^2}{n_1 + n_2 + \ldots n_k}$$

$$+\frac{n_1(m-m_1)^2+n_2(m-m_2)^2+\ldots.n_k(m-m_k)^2}{n_1+n_2+\ldots+n_k}$$

Ex. 2-22. *The standard deviations of two sets containing n_1 and n_2 numbers are σ_1 and σ_2 respectively deviations being measured from their respective means m_1 and m_2. If the two sets are grouped together as one set of (n_1+n_2) members, show that the s.d. σ of this set measured from its mean is given by*

$$\sigma^2=\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2}+\frac{n_1n_2}{(n_1+n_2)^2}(m_1-m_2)^2$$

Sol. Let $x_1, x_2, \ldots. x_{n_1}$ and $y_1, y_2, \ldots. y_{n_2}$ *be the members of two sets. Then by def.*

$$m_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i \text{ and } m_2=\frac{1}{n_2}\sum_{j=1}^{n_2} y_j$$

Let m be the mean of the grouped set. Then

$$m = \frac{1}{n_1+n_2}\left\{\sum_{i=1}^{n_1} x_i+\sum_{j=1}^{n_2} y_j\right\}=\frac{n_1m_1+n_2m_2}{n_1+n_2}$$

Now $$\sigma^2 = \frac{1}{n_1+n_2}\left\{\sum_{i=1}^{n_1}(x_i-m)^2+\sum_{j=1}^{n_2}(y_j-m)^2\right\}$$

$$=\frac{1}{n_1+n_2}\left\{\sum_{i=1}^{n}(x_i-m_1+m_1-m)^2+\sum_{j=1}^{n}(y_j-m_2+m_2-m)^2\right\}$$

$$=\frac{1}{n_1+n_2}\left[\left\{\sum_{i=1}^{n_1}(x_i-m_1)^2+2(m_1-m)\sum_{i=1}^{n_1}(x_i-m_1)+\sum_{i=1}^{n_1}(m_1-m)^2\right\}+\right.$$

$$\left.\left\{\sum_{j=1}^{n_1}(y_j-m_2)^2+2(m_2-m)\sum_{j=1}^{n_2}(y_j-m_2)+\sum_{j=1}^{n_2}(m_2-m)^2\right\}\right]$$

$$=\frac{1}{n_1+n_2}\left\{n_1\sigma_1{}^2+n_1(m_1-m)^2+n_2\sigma_2{}^2+(m_2-m)^2n_2\right\}$$

$$=\frac{n_1\sigma_1{}^2+n_2\sigma_2{}^2}{n_1+n_2}+\frac{n_1(m_1-m)^2+n_2(m_2-m)^2}{n_1+n_2}$$

$$\left(\because \sigma_1{}^2=\frac{1}{n_1}\sum_{i=1}^{n_1}(x_i-m_1)^2 \text{ etc.}\right)$$

Now $(m_1 -$

and $(m_2 -$

$\therefore$

Ex. 2-23. *An analysis of th belonging to the same industry,*

No. of wage earners
Average monthly wage
Variance of the dist. of wag
(a) Which firm A or B pays
(b) In which firm A or B is
(c) What are the measures o wages of all the workers in the f

Sol. (*a*) Firm *A* pays = (52·
Firm *B* pays = (47·
$\therefore$ Firm *B* pays more as mor
(*b*) Co-efficient of variation

Co-efficient of variation for

$\therefore$ Firm *B* has greater variat

(*c*) Average monthly wage a the firm *A* and *B* taken together, by Firms *A* and *B* together.

Let m and σ be the *A.M.* an

Then

and

$$\frac{{}_2(m-m_2)^2+....n_k(m-m_k)^2}{{}_1+n_2+...+n_k}$$

ing n_1 and n_2 numbers are σ_1 respective means m_1 and m_2. If members, show that the s.d. σ of

$$\frac{{}_1n_2}{+n_2)^2}(m_1-m_2)^2$$

ers of two sets. Then by def.

$$\frac{1}{n_2}\sum_{j=1}^{n_2} y_j$$

$$y_j\Big\} = \frac{n_1m_1+n_2m_2}{n_1+n_2}$$

$$)^2+\sum_{j=1}^{n_2}(y_j-m)^2\Big\}$$

$$-m_2+m_2-m)^2\Big\}$$

$$)+\sum_{i=1}^{n_1}(m_1-m)^2\Big\}+$$

$$+\sum_{j=1}^{n_2}(m_2-m)^2\Big\}\Big]$$

$$m)^2+n_2\sigma_2^2+(m_2-m)^2n_2\Big\}$$

$$\frac{-m)^2+n_2(m_2-m)^2}{n_1+n_2}$$

$$\sigma_1^2=\frac{1}{n_1}\sum_{i=1}^{n_1}(x_i-m_1)^2 \text{ etc.}\Big)$$

Now $$(m_1-m)^2 = \frac{n_2^2(m_1-m_2)^2}{(n_1+n_2)^2}$$

and $$(m_2-m)^2 = \frac{n_1^2(m_1-m_2)^2}{(n_1+n_2)^2}$$

$\therefore$ $$\sigma^2 = \frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2}+\frac{n_1n_2(m_1-m_2)^2}{(n_1+n_2)^2}.$$

Ex. 2-23. *An analysis of the monthly wages paid to workers in two firms A and B, belonging to the same industry, gives the following results :*

	Firm A	*Firm B*
No. of wage earners	*586*	*648*
Average monthly wage	*Rs. 52·5*	*Rs. 47·5*
Variance of the dist. of wages	*100*	*121*

(a) Which firm A or B pays out the larger amount as monthly wages ?

(b) In which firm A or B is there greater variability in individual wages ?

(c) What are the measures of (i) average monthly wage and the variability in individual wages of all the workers in the firms A and B taken together ?

Sol. (*a*) Firm *A* pays = (52·5) (586) = Rs. 30765 monthly
Firm *B* pays = (47·5) (648) = Rs. 30780 monthly.

$\therefore$ Firm *B* pays more as monthly wages.

(*b*) Co-efficient of variation for Firm *A*

$$= \left(\frac{\sqrt{100}}{52\cdot5}\right)(100) = \frac{1000}{52\cdot5} = 19\cdot05\%$$

Co-efficient of variation for firm *B*

$$= \left(\frac{\sqrt{121}}{47\cdot5}\right)(100) = \frac{1100}{47\cdot5} = 23\cdot16\%$$

$\therefore$ Firm *B* has greater variability in the individual wages.

(*c*) Average monthly wage and the variability in individual wages of all the workers in the firm *A* and *B* taken together, are the A.M. and co-efficient of variation of the wages paid by Firms *A* and *B* together.

Let m and σ be the *A.M.* and s.d. of the wages paid by *A* and *B* together.

Then $$m = \frac{(586)(52\cdot6)+(648)(47\cdot5)}{586+648} = 49\cdot87$$

and $$\sigma^2 = \frac{(586)(100)+(648)(121)}{1234}$$

$$+\frac{(586)(49\cdot87-52\cdot5)^2+648(49\cdot87-47\cdot5)^2}{1234}$$

$$= 117\cdot26$$

$\therefore$ Co-efficient of variation for firms A and B taken together

$$= \frac{\sqrt{117 \cdot 26}}{49 \cdot 87} \times 100 = \frac{1082 \cdot 87}{49 \cdot 87}$$

$$= 21 \cdot 7\%.$$

Ex. 2-24. *The first of two samples has 100 items with mean 15 and s.d. 3. If the whole group has 250 items with mean 15·6 and s.d.* $\sqrt{13 \cdot 44}$, *find the s.d. of the second group.*

Sol. Let m_2 be the mean of second group

Then $$15 \cdot 6 = \frac{(100)(15) + (150)m_2}{250}$$

$$\therefore \quad m_2 = \frac{(250)(15 \cdot 6) - 100(15)}{150} = 16$$

Let σ be the standard deviation of second group. Then

$$\therefore \quad 13 \cdot 44 = \frac{[(100)(9) + (150)\sigma^2] + 100(0 \cdot 6)^2 + 150(0 \cdot 4)^2}{250}$$

$$\therefore \quad \sigma = 4.$$

Ex. 2-25. *The mean and s.d. of 63 children on an arithmetic test are respectively 27·6 and 7·1. To them are added a new group of 26 who have had less training and whose mean is 19·2 and s.d. 6·2. How will the values of the combined group differ from those of the original 63 children as to the following (i) the mean (ii) the s.d.*

Sol. Mean m and *s.d.* σ of the combined group are given by

(*i*) $$m = \frac{(63)(27 \cdot 6) + (26)(19 \cdot 2)}{63 + 26} = 25 \cdot 1$$

$\therefore$ The A.M. is decreased by $27 \cdot 6 - 25 \cdot 1 = 2 \cdot 5$

(*ii*) $$\sigma^2 = \frac{(63)(7 \cdot 1)^2 + (26)(6 \cdot 2)^2}{63 + 26} + \frac{63(25 \cdot 1 - 27 \cdot 6)^2 + 26(25 \cdot 1 - 19 \cdot 2)^2}{63 + 26}$$

$\therefore$ $\sigma = 7 \cdot 8$ (approx)

$\therefore$ The s.d. is increased by $7 \cdot 8 - 7 \cdot 1 = 0 \cdot 7$ (approx.)

Ex. 2-26. *A distribution consists of three components with frequencies of 200, 250 and 300 having means 25, 10 and 15 and s.d. of 3, 4 and 5 respectively. Show that the mean of the combined distribution is 16 and s.d. 7·2 approximately.*

Sol. Let m and σ be the mean and *s.d.* of the combined distribution.

Then $$m = \frac{(25)(200) + (10)(250) + (15)(300)}{200 + 250 + 300}$$

$$= \frac{5000 + 2500 + 4500}{750} = \frac{12000}{750} = 16$$

and $$\sigma^2 = \frac{(200)(3^2) + (250)(4^2) + (300)(5^2)}{200 + 250 + 300}$$

$$+ \frac{200\{16 - 25\}^2 + 250\{16 - 10\}^2 + 300\{16 - 15\}^2}{200 + 250 + 300}$$

$\therefore$ c

2.4. Moments. *The rth momen*

$\mu_r'(a)$

If 'a' is A.M., rth moment abo

Factorial Moments. *Factoria*

$\mu'_{(r)}$

where $x^{(r}$

Absolute Moments. *Absolute*

to be

Pearson's β *and* γ *Co-efficient*

$\beta_1 =$

$\gamma_1 =$

Moments about mean in terms

$\mu_r = \mu_r' - {}^r c_1 \mu'_{r-1}$

Shappard's Corrections to Mo

No correction applied to odd

μ_3 (corrected) $= \mu_3$ and μ_2

μ_4 (corrected) $= \mu_4 - \frac{1}{2}h^2\mu$

2.4-1. *Moments about mean a*

The transformation correspon

$\therefore$

$\therefore$

where $\bar{x}$ and $\bar{u}$ are A.Ms.

ogether

$$: \frac{1082 \cdot 87}{49 \cdot 87}$$

ı mean 15 and s.d. 3. If the whole
nd the s.d. of the second group.

m_2

$$)(15) = 16$$

:n

$$\frac{\text{5}^2] + 100(0 \cdot 6)^2 + 150(0 \cdot 4)^2}{250}$$

thmetic test are respectively 27·6
'ad less training and whose mean
:d group differ from those of the
he s.d.

ven by

$$\frac{(19 \cdot 2)}{} = 25 \cdot 1$$

$$\frac{^2 + 26(25 \cdot 1 - 19 \cdot 2)^2}{3 + 26}$$

with frequencies of 200, 250 and
spectively. Show that the mean of
y.

d distribution.

$$\frac{250) + (15)(300)}{) + 300}$$

$$\frac{00}{} = \frac{12000}{750} = 16$$

$$\frac{(4^2) + (300)(5^2)}{0 + 300}$$

$$\frac{250\{16 - 10\}^2 + 300\{16 - 15\}^2}{)0 + 250 + 300}$$

$$= \frac{(1800 + 4000 + 7500) + (16200 + 9000 + 300)}{750}$$

$$= \frac{38800}{750} = 51 \cdot 73$$

$$\therefore \qquad \sigma = \sqrt{51 \cdot 73} = 7 \cdot 2 \text{ (approx.)}$$

2.4. Moments. *The rth moment about the point 'a' is defined by*

$$\mu_r{}'(a) = \frac{1}{N} \Sigma f (x - a)^r$$

If 'a' is A.M., rth moment about 'a' is denoted by μ_r.

Factorial Moments. *Factorial moment of order 'r' about the origin is defined by*

$$\mu'_{(r)} = \frac{1}{N} \Sigma f x^{(r)}$$

where
$$x^{(r)} = x(x-1)\ldots\ldots(x-r+1)$$

Absolute Moments. *Absolute moment of order r about an arbitrary point 'a' is defined to be*

$$\frac{1}{N} \Sigma f |x - a|^r$$

Pearson's β *and* γ *Co-efficients. These co-efficients are defined by*

$$\beta_1 = \frac{\mu_3{}^2}{\mu_2{}^3}, \qquad \beta_2 = \frac{\mu_4}{\mu_2{}^2}$$

$$\gamma_1 = \sqrt{\beta_1} \quad \text{and} \quad \gamma_2 = \beta_2 - 3$$

Moments about mean in terms of moments about any other point are given by

$$\mu_r = \mu'_r - {}^r c_1 \mu'_{r-1} \mu'_1 + {}^r c_2 \mu_{r-2} \{\mu'_1\}^2 + \ldots + (-1)^r \, {}^r c_r \{\mu'_1\}^r$$

Shappard's Corrections to Moments of Grouped Frequency Distribution.

No correction applied to odd order moment *i.e.*, $\mu_1{}'$ (corrected) $= \mu_1{}'$

$$\mu_3 \text{ (corrected)} = \mu_3 \text{ and } \mu_2 \text{ (corrected)} = \mu_2 - \frac{h^2}{12}$$

$$\mu_4 \text{ (corrected)} = \mu_4 - \frac{1}{2} h^2 \mu_2 + \frac{7}{240} h^4.$$

2.4-1. *Moments about mean are independent of origin but not of scale.*

The transformation corresponding to change in origin and scale is $u = \dfrac{x - a}{h}$.

$$\therefore \qquad x = a + uh$$

$$\therefore \qquad \bar{x} = a + \bar{u} h$$

where $\bar{x}$ and $\bar{u}$ are A.Ms.

$$\therefore \ \mu_r \text{ of } x = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r = h^r \cdot \frac{1}{N}\sum_{i=1}^{n} f_i(u_i - \bar{u})^r = h^r \cdot (\mu_r \text{ of } u)$$

2.4-2. *Expression of rth moment about mean in terms of various moments about an arbitrary pt. 'a'.*

Let $\bar{x}$ be the A.M. Then

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r = \frac{1}{N}\sum_{i=1}^{n} f_i[(x_i - a) + (a - \bar{x})]^r$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^r + {}^r c_1 (x_i - a)^{r-1}(a - \bar{x})$$

$$+ {}^r c_2 (x_i - a)^{r-2}(a - \bar{x})^2 + \ldots + {}^r c_r (a - \bar{x})^r]$$

$$= \mu_r'(a) + {}^r c_1\, \mu_{r-1}'(a)(a - \bar{x}) + {}^r c_2\, \mu_{r-2}'(a - \bar{x})^2 + \ldots + {}^r c_r (a - \bar{x})^r$$

Now $a - \bar{x} = -\dfrac{1}{N}\displaystyle\sum_{i=1}^{n} f_i(x_i - a) = -\mu_1'(a)$

$$\therefore \ \mu_r = \mu_r'(a) - {}^r c_1\, \mu_{r-1}'(a)\mu_1'(a) + {}^r c_2\, \mu_{r-2}'(a)\{\mu_1'(a)\}^2 + \ldots + (-1)^r\ {}^r c_r\ \{\mu_1'(a)\}^r$$

$$= \sum_{j=0}^{r} {}^r c_j\, \mu_{r-j}' \{-\mu_1'\}^j$$

where μ_r' denotes $\mu_r'(a)$.

2.4-3. *Expression of rth moment about a pt 'a' in terms of various moments about mean.*

Let μ_r' (a) be written as μ_r'.

Then
$$\mu_r' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^r$$

$$= \frac{1}{N}\sum_{i} f_i\{(x_i - \bar{x}) + (\bar{x} - a)\}^r$$

$$= \frac{1}{N}\sum_{i} f_i\{(x_i - \bar{x}) + \mu_1'\}^r$$

$$= \frac{1}{N}\sum f_i\{(x_i - \bar{x})^r + {}^r c_1 (x_i - \bar{x})^{r-1}\mu_1' + {}^r c_2 (x_i - \bar{x})^{r-2}\{\mu_1'\}^2 + \ldots + {}^r c_r\ \{\mu_1'\}^r$$

$$= \mu_r + r_{c_1}\, \mu_{r-1}\, \mu_1' + r_{c_2}\, \mu_{r-2}\{\mu_1'\}^2 \ldots\ldots + {}^r c_r \{\mu_1'\}^r$$

$$= \sum_{j=0}^{r} r_{c_j}\ \mu_{r-j}\{\mu_1'\}^j.$$

2.4-4. *Expression of rth moment other pt. 'b'.*

$$\mu_r'(a) = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^r = \frac{1}{N}$$

$$= \mu_r'(b) + {}^r c_1 (b - a)\mu_{r-1}'($$

Ex. 2-27. *Calculate the first fou calculate β_1 and β_2.*

x values in cm, are the mid-poin

x : 2·0 2·5

f : 5 38

Sol.

Variable (x)	*f*	$d = \frac{x - 3}{0 \cdot 5}$
2·0	5	−3
2·5	38	−2
3·0	65	−1
3·5	92	0
4·0	70	1
4·5	40	2
5·0	0	3
	310	

$\therefore$ $\mu_1'(3 \cdot 5)$

$\mu_2'(3 \cdot 5)$

$\mu_3'(3 \cdot 5)$

2.4-4. *Expression of rth moment about a pt. 'a' in terms of various moments about any other pt. 'b'.*

$$\mu_r'(a) = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - a)^r = \frac{1}{N}\sum_{i=1}^{n} f_i\{(x_i - b) + (b - a)\}^r$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i[(x_i - b)^r + {}^rc_1\,(x_i - b)^{r-1}(b - a)$$

$$+ {}^rc_2\,(x_i - b)^{r-2}(b - a)^2 + \ldots + {}^rc_r\,(b - a)^r]$$

$$= \mu_r'(b) + {}^rc_1(b-a)\mu_{r-1}'(b) + {}^rc_2(b-a)^2\,\mu_{r-2}'(b) + \ldots + {}^rc_r(b-a)^r.$$

Ex. 2-27. *Calculate the first four moments about the mean of the following dist, also calculate* β_1 *and* β_2.

x values in cm, are the mid-points of intervals :

x :	2·0	2·5	3·0	3·5	4·0	4·5	5·0
f :	5	38	65	92	70	40	0

Sol.

Variable (x)	f	$d = \frac{x - 3\cdot 5}{0\cdot 5}$	$d.f$	$d^2.f$	$d^3.f$	$d^4.f$
2·0	5	–3	–15	45	–135	405
2·5	38	–2	–76	152	–304	608
3·0	65	–1	–65	65	–65	65
3·5	92	0	0	0	0	0
4·0	70	1	70	70	70	70
4·5	40	2	80	160	320	640
5·0	0	3	0	0	0	0
	310		–6	492	–114	1788

$$\therefore \quad \mu_1'(3\cdot 5) = \frac{\Sigma d\cdot f}{N} \times (0\cdot 5) = -\frac{6}{310} \times (0\cdot 5)$$

$$= -\frac{3}{310} = -0\cdot 01$$

$$\mu_2'(3\cdot 5) = \frac{\Sigma d^2\cdot f}{N}(0\cdot 5)^2 = \frac{492}{310}(0\cdot 25)$$

$$= \frac{123}{310} = 0\cdot 397.$$

$$\mu_3'(3\cdot 5) = \frac{\Sigma d^3 f}{N}(0\cdot 5)^3$$

$$= -\frac{114}{310}(0\cdot125) = -0\cdot046$$

$$\mu_4{'}(3\cdot5) = \frac{\Sigma d^4 \cdot f}{N} \times (0\cdot5)^4 = \frac{1788}{310}(0\cdot0625)$$

$$= \frac{1788}{310} \times \frac{1}{16} = \frac{447}{1240} = 0\cdot360.$$

Moments about the A.M. are :

$$\mu_1 = 0$$

$$\mu_2 = \mu_2'(3\cdot5) - [\mu_1'(3\cdot5)]^2$$

$$= 0\cdot397 - 0\cdot0001 = 0\cdot3969 \simeq 0\cdot40$$

$$\mu_3 = \mu_3{'}(3\cdot5) - 3\mu_2'(3\cdot5)\,\mu_1'(3\cdot5)] + 2[\mu_1'(3\cdot5)]^3$$

$$= -0\cdot046 - 3(0\cdot397)(-0\cdot01) + 2(-0\cdot01)^3$$

$$= -0\cdot034 = -0\cdot03 \text{ (approx.)}$$

$$\mu_4 = \mu_4{'}(3\cdot5) - 4\mu_3{'}(3\cdot5)\,\mu_1'(3\cdot5) + 6\mu_2'(3\cdot5)[\mu_1'(3\cdot5)]^2 - 3[\mu_1'(3\cdot5)]^4$$

$$= 0\cdot360 - 4(-0\cdot046)(-0\cdot01) + 6(0\cdot397)(-0\cdot01)^2 - 3(-0\cdot01)^4$$

$$= 0\cdot358 = 0\cdot36 \text{ (approx.)}.$$

$$\beta_1 = \frac{\mu_3{}^2}{\mu_2{}^3} = \frac{(-0\cdot034)^2}{(0\cdot397)^3} = 0\cdot02$$

$$\beta_2 = \frac{\mu_4}{\mu_2{}^2} = \frac{0\cdot358}{(0\cdot397)^2} = 2\cdot27$$

Ex. 2-28. *The first four moments of a distribution about the value 4 are –1·5, 17, –30, 108. Calculate the moments about the mean.*

Sol. Moment about the mean are

$$\mu_1 = 0$$

$$\mu_2 = \mu_2{'} - \mu_1{'}^2 = 17 - (-1\cdot5)^2$$

$$= 17 - 2\cdot25 = 14\cdot75$$

$$\mu_3 = \mu_3{'} - 3\mu_2{'}\,\mu_1{'} + 2\mu_1{'}^3$$

$$= -30 - 3(17)(-1\cdot5) + 2(-1\cdot5)^3$$

$$= -30 + 76\cdot5 - 6\cdot75$$

$$= 39\cdot75$$

$$\mu_4 = \mu_4' - 4\mu_3'\,\mu_1' + 6\mu_2'\,\mu_1{'}^2 - 3\mu_1{'}^4$$

$$= 108 - 4(-30)(-1\cdot5) + 6(17)(-1\cdot5)^2 - 3(-1\cdot5)^4$$

$$= 108 - 180 + 229\cdot5 - 15\cdot1875$$

$$= 142\cdot3125 \simeq 142\cdot3$$

Ex. 2-29. *The first four moments*
2, 20, 40 and 50. Obtain as far as poss
the basis of the information given.

Sol. The first four moments abou

$\mu_1 =$

$\mu_2 =$

$\mu_3 =$

$\mu_4 =$

$A.M. = \mu_1{'}(0)$

$\beta_1 = \frac{\mu_3{}^2}{\mu_2{}^3}$

$\beta_2 = \frac{\mu_4}{\mu_2{}^2}$

$\gamma_1 = \sqrt{\beta_1}$

$\gamma_2 = \beta_2 -$

Ex. 2-30. *The first three moment*
1, 16, –40. Find as far as you can, the
information given.

Sol. A.M.

$= -0\cdot046$

$4 = \frac{1788}{310}(0\cdot0625)$

$\frac{47}{40} = 0\cdot360.$

$3\cdot5)]^2$

$1 = 0\cdot3969 \simeq 0\cdot40$

$(3\cdot5)\mu_1'(3\cdot5)] + 2[\mu_1'(3\cdot5)]^3$

$397)(-0\cdot01) + 2(-0\cdot01)^3$

3 (approx.)

$_1'(3\cdot5)]^2 - 3[\mu_1'(3\cdot5)]^4$

$01)^2 - 3(-0\cdot01)^4$

$)^2 \; _3 = 0\cdot02$

$= 2\cdot27$

bout the value 4 are –1·5, 17, –30,

$-(-1\cdot5)^2$

5

$2\mu_1'^3$

$5) + 2(-1\cdot5)^3$

75

$\mu_2'\mu_1'^2 - 3\mu_1'^4$

$\cdot5) + 6(17)(-1\cdot5)^2 - 3(-1\cdot5)^4$

$5 - 15\cdot1875$

$\cdot3$

Ex. 2-29. *The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50. Obtain as far as possible the various characteristics of this distribution on the basis of the information given.*

Sol. The first four moments about the mean are :

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 20 - 4 = 16$$

$$\begin{aligned}\mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3\\ &= 40 - 3(20)(2) + 2(2)^3\\ &= 40 - 120 + 16\\ &= -64\end{aligned}$$

$$\begin{aligned}\mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4\\ &= 50 - 4(40)(2) + 6(20)(2)^2 - 3(2)^4\\ &= 50 - 320 + 480 - 48\\ &= 530 - 368 = 162\end{aligned}$$

$$\begin{aligned}A.M. = \mu_1'(0) &= \frac{1}{N}\Sigma fx\\ &= \frac{1}{N}\Sigma f[(x-5)+5]\\ &= \frac{1}{N}\Sigma f(x-5) + 5\\ &= \mu_1'(5) + 5 = 2 + 5 = 7.\end{aligned}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-64)^2}{(16)^3} = 1$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{162}{(16)^2} = \frac{81}{128} = 0\cdot63.$$

$$\gamma_1 = \sqrt{\beta_1} = 1$$

$$\gamma_2 = \beta_2 - 3 = 0\cdot63 - 3 = -2\cdot37.$$

Ex. 2-30. *The first three moments of a distribution about the value 2 of the variable are 1, 16, –40. Find as far as you can, the various characteristics of this dist on the basis of the information given.*

Sol.

$$\begin{aligned}\text{A.M.} &= \frac{1}{N}\Sigma fx = \frac{1}{N}\Sigma f(x-2)+2]\\ &= \frac{1}{N}\Sigma f(x-2)+2\end{aligned}$$

$$= \mu_1'(2) + 2$$
$$= 1 + 2 = 3.$$

The first three moments about the A.M. are given by

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 16 - 1 = 15$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$
$$= -40 - 3(16)(1) + 2(1)^3$$
$$= -40 - 48 + 2 = -86$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-86)^2}{(15)^3} = 2\cdot19$$

$$\therefore \quad \gamma_1 = \sqrt{\beta_1} = \sqrt{2\cdot19} = 1\cdot48.$$

Ex. 2-31. *The first four moments of a distribution are 1, 4, 10 and 46 respectively. Compute the first four central moments and beta constants. Comment upon the nature of the dist.*

Sol. The first four central moments are given by :

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 4 - 1 = 3$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$
$$= 10 - 3(4)(1) + 2(1)^3$$
$$= 10 - 12 + 2 = 0$$

$$\mu_4 = \mu_4 - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$
$$= 46 - 4(10)(1) + 6(4)(1)^2 - 3(1)^4$$
$$= 46 - 40 + 24 - 3$$
$$= 27$$

$$\therefore \quad \beta_1 = 0,\ \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{27}{9} = 3.$$

Since $\beta_1 = 0, \beta_2 = 3,$ the distribution must be normal.

Ex. 2-32. *For a distribution of 250 heights, calculations showed that the mean, standard deviation, β_1 and β_2 were 54 inches, 3 inches, 0 and 3 inches respectively. It was, however, discovered on checking that the two items 64 and 50 in the original data were wrongly written in place of correct values 62 and 52 inches respectively. Calculate the correct frequency constants.*

Sol. Let x be the variable and N be the total frequency. Then $N = 250$.

Then $\Sigma x = (250)(54)$

∴ Corrected value of Σx

∴ Corrected A.M

Varianc

or $\frac{1}{N}\Sigma(x-\bar{x})^2$

or $\Sigma(x-\bar{x})^2 = (250)(9)$

∴ Corrected $\Sigma(x-\bar{x})^2 = 225$

∴ Corrected varianc

Now $\beta_1 = ($

∴ μ

∴ $\Sigma(x-\bar{x})$

∴ Corrected $\Sigma(x-\bar{x})^3$

$= 0 - (64 - 54)$

$= 0 - 1000 + 6$

$= -432$

∴ Corrected $\mu_3 = \frac{-432}{250} = -1$

∴ Corrected $\beta_1 = \frac{\text{(corrected}}{\text{(corrected}}$

$= \frac{(-1\cdot728)^2}{(8\cdot808)^3}$

$= 0\cdot004$

β

∴ μ

∴ $\Sigma(x-\bar{x}$

∴ Corrected $\Sigma(x-\bar{x})^4 = 60$

∴ Corrected μ

y

$-1 = 15$

$2\mu_1'^3$

$2(1)^3$

$\cdot 86$

$\cdot 19$

$\cdot 48.$

are 1, 4, 10 and 46 respectively.
nts. Comment upon the nature of

$= 3$

$2\mu_1'^3$

$)^3$

$_2\, \mu_1'^2 - 3\mu_1'^4$

$4)(1)^2 - 3(1)^4$

$= 3.$

s showed that the mean, standard
hes respectively. It was, however,
the original data were wrongly
pectively. Calculate the correct

. Then $N = 250$.

$\therefore$ Corrected value of $\Sigma x = (250)(54) - 64 - 50 + 62 + 52 = (250)(54)$

$\therefore$ Corrected A.M. $= 54$.

Variance $= 9$.

or $$\frac{1}{N}\Sigma(x-\bar{x})^2 = 9$$

or $$\Sigma(x-\bar{x})^2 = (250)(9) = 2250$$

$$\therefore \text{Corrected } \Sigma(x-\bar{x})^2 = 2250 - (64-54)^2 - (50-54)^2 + (62-54)^2 + (52-54)^2$$
$$= 2250 - 100 - 16 + 64 + 4$$
$$= 2202$$

$$\therefore \quad \text{Corrected variance} = \frac{2202}{250} = 8\cdot 808$$

Now $$\beta_1 = 0 = \frac{\mu_3^2}{\mu_2^3}$$

$$\therefore \quad \mu_3 = 0$$

$$\therefore \quad \Sigma(x-\bar{x})^3 = 0$$

$\therefore$ Corrected $\Sigma(x-\bar{x})^3$

$$= 0 - (64-54)^3 - (50-54)^3 + (62-54)^3 + (52-54)^3$$
$$= 0 - 1000 + 64 + 512 - 8$$
$$= -432$$

$$\therefore \text{Corrected } \mu_3 = \frac{-432}{250} = -1\cdot 728$$

$$\therefore \text{Corrected } \beta_1 = \frac{(\text{corrected } \mu_3)^2}{(\text{corrected } \mu_2)^3}$$
$$= \frac{(-1\cdot 728)^2}{(8\cdot 808)^3}$$
$$= 0\cdot 004$$

$$\beta_2 = 3 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{(9)^2}$$

$$\therefore \quad \mu_4 = 243$$

$$\therefore \quad \Sigma(x-\bar{x})^4 = (243)(250) = 60750$$

$$\therefore \text{Corrected } \Sigma(x-\bar{x})^4 = 60750 - (64-54)^4 - (50-54)^4 + (62-54)^4 + (52-54)^4$$
$$= 60750 - 10000 - 256 + 4096 + 16$$
$$= 54606$$

$$\therefore \quad \text{Corrected } \mu_4 = \frac{54606}{250} = 218\cdot 424$$

$$\therefore \qquad \text{Corrected } \beta_2 = \frac{\text{corrected } \mu_4}{(\text{corrected } \mu_2)^2} = \frac{218\cdot424}{(8\cdot808)^2}$$

$$= \frac{218\cdot4}{(8\cdot808)^2} \text{ (approx.)}$$

$$= 2\cdot815.$$

Ex. 2-33. *Second, third and fourth central moments of a variable characteristics are 19·67, 29·26 and 866·0 respectively. Calculate the beta constants correct to three decimal places.*

Sol.
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(29\cdot26)^2}{(19\cdot67)^3} = 0\cdot113 \text{ (approx.)}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{866\cdot0}{(19\cdot67)^2} = 2\cdot238$$

2.4-5. *For a symmetrical distribution all moments about the mean of odd order are zero.*

For a symmetrical distribution the frequencies are symmetrically distributed about the mean *i.e.*, the values equidistant from the mean have equal frequencies. Let x be the variable and $\bar{x}$ its A.M.

Let $y = x - \bar{x}$.

Let x_1, x_2 be the values of x equidistant from $\bar{x}$.

Then the quantities $(x_1 - \bar{x})$ and $(x_2 - \bar{x})$ are equal in magnitude but opposite in signs. Let these quantities by y_1 and $-y_1$. Then since the distribution is symmetrical the values $-y_1$ and y_1 of y have same frequencies f_1 each. Let other values of y be $-y_2; y_2, -y_3, y_3$ and so on. Let $f_2, f_3 \ldots$ be the frequencies for $-y_2; y_2, -y_3; y_3 \ldots\ldots\ldots$ Let N be the total frequency.

Now by def.,

$$\mu_{2r+1} = \frac{1}{N}\Sigma f(x-\bar{x})^{2r+1}$$

$$= \frac{1}{N}\Sigma f y^{2r+1}$$

$$= \frac{1}{N}\{f_1(y_1^{2r+1} - y_1^{2r+1}) + f_2(y_2^{2r+1} - y_2^{2r+1}) + \ldots\} = 0.$$

2.4-6. *Show that for a discrete dist* $\beta_2 > 1$.

By def $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

$\therefore\ \beta_2 > 1$ if $\mu_4 > \mu_2^2$

i.e., $\dfrac{1}{N}\displaystyle\sum_{i=1}^{n} f_i(x_i - \bar{x})^4 > \left\{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i-\bar{x})^2\right\}^2$ where $\bar{x}$ = A.M.

i.e., $N\displaystyle\sum_{i=1}^{n} f_i y_i^2 > \left(\sum_{i=1}^{n} f_i y_i\right)^2$

i.e., $(f_1 + f_2 + \ldots + f_n)(f_1 y_1^2 +$

i.e., $f_1(f_1 y_1^2 + f_2 y_2^2 + \ldots\ldots\ldots +$

$+ \ldots + f_n(f_1 y_1^2 + f_2 y_2^2 + \ldots$

i.e., $f_1 f_2 (y_1 - y_2)^2 + \ldots\ldots > 0$

which is true as each term on the le

$\therefore$

Theorem 2.4-7. *Show that fo*

Proof.

By def,

$\therefore\ \beta_2 > \beta_1$ *if*

i.e.,

i.e., $\left\{\dfrac{1}{N}\displaystyle\sum_{i=1}^{n} f_i(x_i-\bar{x})^4\right\}.\left\{\dfrac{1}{N}\right.$

i.e., $\left(\displaystyle\sum_{i=1}^{n} f_i y_i^4\right)\left(\sum_{i=1}^{n} f_i y_i^2\right) >$

i.e., $\left(f_1 y_1^4 + f_2 y_2^4 + \ldots\ldots\ldots + f_n\right.$

i.e., $f_1 y_1^4\left(f_1 y_1^2 + f_2 y_2^2 + \ldots\right.$

i.e., $f_1 f_2 y_1^2 y_2^2 (y_1 - y_2)^2 + \ldots$

which is true as each term on left

$\therefore\ \beta_2 > \beta_1$.

$$= \frac{218 \cdot 424}{(8 \cdot 808)^2}$$

:.)

of a variable characteristics are
onstants correct to three decimal

3 (approx.)

3

ibout the mean of odd order are

nmetrically distributed about the
frequencies. Let x be the variable

magnitude but opposite in signs.

oution is symmetrical the values

values of y be $-y_2; y_2, -y_3, y_3$

$y_3; y_3$.......... Let N be the total

= A.M.

$$\text{i.e., } N\sum_{i=1}^{n} f_i y_i^2 > \left(\sum_{i=1}^{n} f_i y_i\right)^2 \quad \text{where } y_i = (x_i - \bar{x})^2$$

$$\text{i.e., } (f_1 + f_2 + \ldots + f_n)(f_1 y_1^2 + f_2 y_2^2 + \ldots + f_n y_n^2) > (f_1 y_1 + f_2 y_2 + \ldots\ldots + f_n y_n)^2$$

$$\text{i.e., } f_1(f_1 y_1^2 + f_2 y_2^2 + \ldots\ldots + f_n y_n^2) + f_2(f_1 y_1^2 + f_2 y_2^2 + \ldots\ldots + f_n y_n^2)$$

$$+ \ldots + f_n(f_1 y_1^2 + f_2 y_2^2 + \ldots + f_n y_n^2) > f_1^2 y_1^2 + f_2^2 y_2^2 + \ldots + f_n^2 y_n^2 + 2 f_1 f_2 y_1 y_2 + \ldots$$

$$\text{i.e., } f_1 f_2 (y_1 - y_2)^2 + \ldots\ldots > 0$$

which is true as each term on the left is positive.

$$\therefore \qquad \beta_2 > 1.$$

Theorem 2.4-7. *Show that for a discrete dist.*

$$\beta_2 > \beta_1$$

Proof.

By def,

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$\therefore \beta_2 > \beta_1$ *if*

$$\frac{\mu_4}{\mu_2^2} > \frac{\mu_3^2}{\mu_2^3}$$

$$\text{i.e., } \qquad \mu_4 \mu_2 > \mu_3^2$$

$$\text{i.e., } \left\{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^4\right\} \cdot \left\{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2\right\} > \left\{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^3\right\}^2$$

$$\text{i.e., } \left(\sum_{i=1}^{n} f_i y_i^4\right)\left(\sum_{i=1}^{n} f_i y_i^2\right) > \left(\sum_{i=1}^{n} f_i y_i^3\right)^2 \quad \text{where } y_i = x_i - \bar{x}$$

$$\text{i.e., } \left(f_1 y_1^4 + f_2 y_2^4 + \ldots\ldots + f_n y_n^4\right)\left(f_i y_1^2 + \ldots\ldots + f_n y_n^2\right)$$

$$> \left\{f_i y_1^3 + f_2 y_2^3 + \ldots\ldots + f_n y_n^3\right\}^2$$

$$\text{i.e., } f_1 y_1^4 \left(f_1 y_1^2 + f_2 y_2^2 + \ldots + f_n y_n^2\right) + f_2 y_2^4 \left(f_1 y_1^2 + f_2 y_2^2 + \ldots + f_n y_n^2\right) \ldots$$

$$> f_1^2 y_1^6 + f_2^2 y_2^6 + \ldots + f_n^2 y_n^6 + 2 f_1 f_2 y_1^3 y_2^3 + \ldots$$

$$\text{i.e., } f_1 f_2 y_1^2 y_2^2 (y_1 - y_2)^2 + \ldots\ldots > 0$$

which is true as each term on left is non-negative.

$$\therefore \beta_2 > \beta_1.$$

2.4-8, *Expression of first four factorial moments in terms of ordinary moments about origin and conversely.*

Factorial moment of order 'r' about the origin is defined by

$$\mu'_{(r)} = \frac{1}{N}\sum_{i=1} f_i x_i^{(r)}$$

where $$x^{(r)} = x(x-1)....(x-r+1)$$

Now $$\mu'_{(1)} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^{(1)} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i = \mu_1'(0)$$

$$\mu'_{(2)} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^{(2)} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i(x_i-1)$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \frac{1}{N}\sum_{i=1}^{n} f_i x_i = \mu_2'(0) - \mu_1'(0)$$

$$\mu'_{(3)} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i(x_i-1)(x_i-2)$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \{x_i^3 - 3x_i^2 + 2x_i\} = \mu_3'(0) - 3\mu_2'(0) + 2\mu_1'(0)$$

$$\mu'_{(4)} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i(x_i-1)(x_i-2)(x_i-3)$$

$$= \mu_4'(0) - 6\mu_3'(0) + 11\mu_2'(0) - 6\mu_1'(0)$$

Conversely : $\mu_1'(0) = \mu_{(1)}'$

$$\mu_{(2)}'(0) = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 = \frac{1}{N}\sum_{i=1}^{n} f_i \{x_i(x_i-1) + x_i\}$$

$$= \mu'_{(2)} + \mu'_{(1)}$$

$$\mu_3'(0) = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^3$$

Let $$x^3 \equiv x(x-1)(x-2) + Ax(x-1) + Bx$$

Put $$x = 1, 2$$

$\therefore$ $$B = 1 \text{ and } 2A + 2B = 8 \text{ or } A=3$$

$\therefore$ $$x^3 \equiv x(x-1)(x-2) + 3x(x-1) + x$$

$\therefore$ $$\mu_3'(0) = \frac{1}{N}\sum_{i=1}^{n} f_i \{x_i(x_i-1)(x_i-2) + 3x_i(x_i-1) + x_i\}$$

=

$\mu_4'(0) =$

Now $x^4 \equiv x(x-1)(x-2)(x$

$\therefore$ $\mu_4'(0) =$

2.4.9 Absolute moment of

and the absolute moment of ord

Evidently mean deviation i

Ex. 2-34. *Define absolute*

where A_r *is the rth absolute m*

Let a and b be any two nu

$\sum_{i=1}^{n} f_i \{a$

i.e., $\sum_{i=1}^{n} f_i \{a$

i.e., $a^2 \frac{1}{N}\sum_{i=1}^{n} f_i |y$

Put $y_i = x_i$

Then $a^2 . \frac{1}{N}\sum_{i=1}^{n} f_i |x$

i.e., $a^2 A_{r-1} + b^2 A_{r+1} + 2$

i.e., $A_{r-1}\left\{a + \frac{A_r}{A_{r-1}} b\right\}^2 +$

rms of ordinary moments about

ed by

$= \mu_1'(0)$

$(x_i - 1)$

$= \mu_2'(0) - \mu_1'(0)$

$= \mu_3'(0) - 3\mu_2'(0) + 2\mu_1'(0)$

$x_i - 3)$

$0) - 6\mu_1'(0)$

$x_i - 1) + x_i\}$

Bx

x

$+ 3x_i(x_i - 1) + x_i\}$

$$= \mu'_{(3)} + 3\mu'_{(2)} + \mu'_{(1)}$$

$$\mu_4'(0) = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^4$$

Now $x^4 \equiv x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x$

$$\therefore \qquad \mu_4'(0) = \mu_{(4)}' + 6\mu_{(3)}' + 7\mu_{(2)}' + \mu_{(1)}'$$

2.4.9 Absolute moment of order r about the origin is defined by

$$\frac{1}{N}\sum_{i=1}^{n} f_i |x_i|^r$$

and the absolute moment of order r about any arbitrary pt 'a' is defined by

$$\frac{1}{N}\sum_{i=1}^{n} f_i |x_i - a|^r$$

Evidently mean deviation is the first order absolute moment.

Ex. 2-34. *Define absolute moments. Show that*

$$A_r^{2r} \le . A^r{}_{r-1} . A^r{}_{r+1}$$

where A_r is the rth absolute moment about point. Deduce that

$$A_r^{\frac{1}{r}} \le A_{r+1}^{\frac{1}{r+1}} \quad r = 1, 2 \ldots\ldots$$

Let a and b be any two numbers. Then

$$\sum_{i=1}^{n} f_i \left\{ a|y_i|^{\frac{r-1}{2}} + b|y_i|^{\frac{r+1}{2}} \right\}^2 \ge 0$$

i.e., $$\sum_{i=1}^{n} f_i \{a^2|y_i|^{r-1} + b^2|y_i|^{r+1} + 2ab|y_i|^r\} \ge 0$$

i.e., $$a^2 \frac{1}{N}\sum_{i=1}^{n} f_i |y_i|^{r-1} + b^2 \cdot \frac{1}{N}\sum_{i=1}^{n} f_i |y_i|^{r+1} + 2ab\frac{1}{N}\sum_{i=1}^{n} f_i |y_i|^r \ge 0$$

Put $\qquad y_i = x_i - \xi.$

Then $$a^2 . \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \xi|^{r-1} + b^2 \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \xi|^{r+1} + 2ab . \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \xi|^r \ge 0$$

i.e., $a^2 A_{r-1} + b^2 A_{r+1} + 2abA_r \ge 0$

i.e., $A_{r-1}\left\{a + \dfrac{A_r}{A_{r-1}} b\right\}^2 + \left\{A_{r+1} - . \dfrac{A_r^2}{A_{r-1}}\right\} b^2 \ge 0 \qquad$ (if $A_{r-1} \neq 0$)

$$\therefore \quad A_{r+1} - \frac{A_r^2}{A_{r-1}} \geq 0 \text{ or } A_{r+1}A_{r-1} \geq A_r^2$$

$$\therefore \qquad A_r^{2r} \leq A_{r+1}^r . A_{r-1}^r .$$

Put $\qquad r = 1, 2, 3 \ldots r$

$$A_1^2 \leq A_2 A_0$$

$$A_2^4 \leq A_3^2 A_1^2$$

$$A_3^6 \leq A_4^3 A_2^3$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$A_r^{2r} \leq A_{r+1}^r \, A_{r-1}^r$$

Multiplying and using $A_0 = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \xi|^0 = 1$

$$A_r^{r+1} \leq A_{r+1}^r \quad \text{for } r = 1, 2 \ldots\ldots$$

$$A_r^{\frac{1}{r}} \leq A_{r+1}^{\frac{1}{r+1}}$$

Ex. 2-35. *Show that if the class interval of a grouped dist is less than one-third of the calculated s.d. Sheppard's adjustment makes a difference of less than* $\frac{1}{2}\%$ *in the estimate of s.d.*

Sol. From Sheppard's correction

$$\mu_2 \text{ (corrected)} = \mu_2 - \frac{h^2}{12}$$

Let $\qquad \mu_2 \text{ (corrected)} = \sigma_1^2$ and $\mu_2 = \sigma^2$

Then
$$\sigma_1 = \left(\sigma^2 - \frac{h^2}{12}\right)^{\frac{1}{2}}$$

$$= \sigma\left(1 - \frac{h^2}{12\sigma^2}\right)^{\frac{1}{2}} = \sigma\left(1 - \frac{h^2}{24\sigma^2} + \ldots\ldots\right)$$

$$\therefore \qquad \sigma - \sigma_1 = \frac{h^2}{24\sigma}$$

Now $\qquad h < \frac{1}{3}\sigma$

$$\therefore \qquad \sigma - \sigma_1 < \frac{\sigma}{216} < \frac{\sigma}{200}$$

$$\therefore \qquad \frac{\sigma - \sigma_1}{\sigma} < \frac{1}{20}$$

2.5. Skewness

For a symmetrical distributic
frequencies is same on both sides
A distribution which is not sy
lack in symmetry. It may be positi

and median lies in between the tw
tail on right) and in negatively sk
Skewness is measured by eitl
Skewn
Skewn
Skewn
For moderate skewed distrib

mean

Co-efficient of Skewness

Bowley's co-efficient of Skew

Karl Pearson's co-efficient c

2nd formula is used when m

In terms of β_1 and β_2,

Coeff. of skewn

This is also zero for symme

Ex. 2-36. *Compute Q.D. an*

Size
4—8
8—12
12—16
16—20
20—24

Sol. We have

$$\therefore \qquad \frac{\sigma - \sigma_1}{\sigma} < \frac{1}{200} = \frac{1}{2}\%.$$

2.5. Skewness

For a symmetrical distribution mean, mode and median coincide. The spread of the frequencies is same on both sides of the mean.

A distribution which is not symmetrical is called **skewed** distribution. Skewness means lack in symmetry. It may be positive or negative. In positively skewed distribution.

$$\text{mode} < \text{mean}$$

and median lies in between the two. Curve is more elongated to the right (*i.e.*, has a longer tail on right) and in negatively skewed distribution reverse is the case.

Skewness is measured by either of the formulae :

$$\text{Skewness} = \text{mean} - \text{mode}$$

$$\text{Skewness} = 3\,(\text{mean} - \text{median})$$

$$\text{Skewness} = Q_3 + Q_1 - 2\,(\text{median})$$

For moderate skewed distributions

$$\text{mean} - \text{median} \cong \frac{1}{3}\,(\text{mean} - \text{mode})$$

Co-efficient of Skewness

Bowley's co-efficient of Skewness

$$= \frac{Q_3 + Q_1 - 2\,Median}{Q_3 - Q_1}$$

Karl Pearson's co-efficient of Skewness

$$= \frac{Mean - Mode}{S.D.}$$

$$= \frac{3\,(Mean - Median)}{S.D.}$$

2nd formula is used when mode is ill-defined.

In terms of β_1 and β_2,

$$\text{Coeff. of skewness} = \frac{\sqrt{\beta_1}\,(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

This is also zero for symmetrical distribution as $\beta_1 = 0$.

Ex. 2-36. *Compute Q.D. and co-efficient of skewness from the following data.*

Size	*Freq.*	*Size*	*Freq.*
4—8	6	24—28	12
8—12	10	28—32	10
12—16	18	32—36	6
16—20	30	36—40	2
20—24	15		

Sol. We have

$$Q_1 = 14{\cdot}5$$

$$Q_3 = 24{\cdot}92$$

Median has $\frac{109}{2} = 54.5$ items below it.

$\therefore$ It lies in 16 – 20

$$\therefore \quad \text{Median} = 16 + \frac{4}{30}(54.5 - 34)$$

$$= 16 + \frac{4}{30}(20.5)$$

$$= 16 + \frac{82}{30} = 16 + 2.73$$

$$= 18.73$$

$\therefore$ Bowley's coefficient of skewness

$$= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{24.92 + 14.5 - 37.46}{24.92 - 14.5}$$

$$= \frac{39.42 - 37.46}{10.42} = \frac{1.96}{10.42}$$

$$= 0.19.$$

Ex. 2-37. *From the data given below calculate Karl Pearson's co-efficient of skewness :*

Marks less than 10	*No. of students 5*
Marks less than 20	*No. of students 12*
Marks less than 30	*No. of students 32*
Marks less than 40	*No. of students 44*
Marks less than 50	*No. of students 50*

Sol.

Calculation of S.D. and A.M.

Class intervals	*Freq. (f)*	*Mid-points*	$X = \frac{x-25}{10}$	fX	fX^2
0—10	5	5	– 2	– 10	20
10—20	7	15	– 1	– 7	7
20—30	20	25	0	0	0
30—40	12	35	1	12	12
40—50	6	45	2	12	24
	50			7	63

$$\text{A.M.} = 25 + 10\left(\frac{7}{50}\right) = 26.4$$

$$\text{S.D.} = 10\sqrt{\frac{63}{50} - \left(\frac{7}{50}\right)^2} = \frac{1}{5}\sqrt{3101} = 11.14.$$

For calculation of mode see next page.

Determination of Model class.

Analysis Table

Class intervals	Given Frequency c. Freq.	Frequency (f) (1)	(2)	(3)	(4)	(5)	(6)	Columns (1)	(2)	(3)	(4)	(5)	(6)	
0—10	5	5	12								✓			1
10—20	12	7		✓ 27	✓ 32					✓	✓	✓		3
20—30	32	✓ 20	✓ 32			✓ 39		✓	✓	✓	✓	✓	✓	6
30—40	44	12		18			✓ 38		✓			✓	✓	3
40—50	50	6											✓	1

$\therefore$ Modal class is 20 – 30

$$\therefore \quad \text{Mode} = 20 + \frac{20-7}{40-7-12} \times 10 = 26{\cdot}19$$

$$\therefore \quad \text{Co-efficient of skewness} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

$$= \frac{26\cdot4 - 26\cdot19}{11\cdot14} = 0\cdot02$$

Ex. 2-38. (*a*) *Find co-efficient of variation if S.D.=3·5, N=10,* $\Sigma x = 145$.

(*b*) *Find the co-efficient of skewness if*

Difference of the two quartiles = 8
Sum of the two quartiles = 22
Median = 10·5

(*c*) *For a series the value of M.D. is 15. Find the most likely value of its Q.D.*

Sol. (*a*) Let $\bar{x}$ be the A.M.

Then
$$\bar{x} = \frac{\Sigma x}{N} = \frac{145}{10} = 14\cdot5$$

$\therefore$ Co-efficient of variation

$$= \left(\frac{\text{S.D.}}{\text{A.M.}}\right)(100)$$

$$= \left(\frac{3\cdot5}{14\cdot5}\right)(100)$$

$$= \frac{3500}{145} = 24\cdot14\%$$

(*b*) Here
$$Q_3 - Q_1 = 8$$
$$Q_3 + Q_1 = 22$$
$$\text{Median} = 10\cdot5$$

$$\therefore \quad \text{Co-efficient of skewness} = \frac{Q_3 + Q_1 - 2\,(\text{Median})}{Q_3 - Q_1}$$

$$= \frac{22 - 21}{8} = \frac{1}{8} = 0\cdot125$$

(*c*) Let σ be the S.D.

Then
$$\text{M.D.} = \frac{4}{5}\sigma$$

and
$$\text{Q.D.} = \frac{2}{3}\sigma$$

$$\therefore \quad \frac{\text{M.D.}}{\text{Q.D.}} = \frac{12}{10} = \frac{6}{5}$$

$$\therefore \quad \text{Q.D.} = \frac{5}{6}(\text{M.D.})$$

$$= \frac{5}{6}(15) = \frac{25}{2} = 12\cdot5$$

$\therefore$ Most likely value of Q.D. =

Ex. 2-39. *Compute quartile de*
following values :

Median = $18\cdot8''$, $Q_1 = 14\cdot6''$, Q

Sol. Q.D.

Co-efficient of skewness

Ex. 2-40. (*a*) *Karl Pearson's co-*
6·5 and mean is 29·6. Find the mode

(*b*) *If the mode of the above dist*

Sol. (*a*) We have

Karl Pearson's co-efficient of sk

$\therefore$ $0\cdot32$

$\therefore$ Mode

Also

Karl Pearson's co-efficient of sl

$\therefore$ Median is given by

$0\cdot32$

$\therefore$ Mediar

(*b*) Mean = 29·6, Mode = 24·8

Karl Pearson's co-efficient of s

$\therefore$ S.D. is given by

$0\cdot32$

i.e. S.D

·02

$5,\ N=10,\ \Sigma x = 145.$

st likely value of its Q.D.

n)

5

$\therefore$ Most likely value of Q.D. = 12·5.

Ex. 2-39. *Compute quartile deviation and the co-efficient of skewness, given the following values :*

Median $= 18 \cdot 8'', Q_1 = 14 \cdot 6'', Q_3 = 25 \cdot 2''$

Sol.

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{25 \cdot 2 - 14 \cdot 6}{2} = 5 \cdot 3$$

Co-efficient of skewness

$$= \frac{Q_3 + Q_1 - 2\,(\text{Median})}{Q_3 - Q_1}$$

$$= \frac{39 \cdot 8 - 37 \cdot 6}{10 \cdot 6} = \frac{2 \cdot 2}{10 \cdot 6} = 0 \cdot 2.$$

Ex. 2-40. (*a*) *Karl Pearson's co-efficient of skewness of a distribution is 0·32. Its s.d. is 6·5 and mean is 29·6. Find the mode and median of the distribution.*

(*b*) *If the mode of the above distribution is 24·8, what will be the standard deviation ?*

Sol. (*a*) We have

Karl Pearson's co-efficient of skewness

$$= \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

$$\therefore \qquad 0 \cdot 32 = \frac{29 \cdot 6 - \text{Mode}}{6 \cdot 5}$$

$$\therefore \qquad \text{Mode} = 29 \cdot 6 - (0 \cdot 32)(6 \cdot 5) = 27 \cdot 52$$

Also

Karl Pearson's co-efficient of skewness

$$= \frac{3\,(\text{Mean} - \text{Median})}{\text{S.D.}}$$

$\therefore$ Median is given by

$$0 \cdot 32 = \frac{3(29 \cdot 6 - \text{Median})}{6 \cdot 5}$$

$$\therefore \qquad \text{Median} = 29 \cdot 6 - \frac{1}{3}(0 \cdot 32)(6 \cdot 5)$$

$$= 29 \cdot 6 - \frac{1}{3}(2 \cdot 08)$$

$$= 29 \cdot 6 - 0 \cdot 69 = 28 \cdot 91.$$

(*b*) Mean = 29·6, Mode = 24·8

Karl Pearson's co-efficient of skewness

$$= 0 \cdot 32$$

$\therefore$ S.D. is given by

$$0 \cdot 32 = \frac{29 \cdot 6 - 24 \cdot 8}{\text{S.D.}}$$

i.e.

$$\text{S.D.} = \frac{4 \cdot 8}{0 \cdot 32} = \frac{480}{32} = 15.$$

2.6. Kurtosis

Let there be two frequency distributions which have same variability as measured by the standard deviation. Their frequency curve may not be equally flat at the top. The flatness of top of a frequency curve is measured relative to that of normal curve. This relative flatness is called **kurtosis.**

Measure of kurtosis tells us the extent to which a frequency curve is more peaked or flat-topped than the normal curve. Kurtosis is measured by β_2. For normal curve, value of β_2 is 3.

For curves which are more peaked than the normal curve, $\beta_2 > 3$. Such curves are called **leptokurtic** and for curves which are more flat-topped than the normal curve, $\beta_2 < 3$. Such curves are called **platykurtic.**

The normal curve itself is called **mesokurtic.**

The greater is the value of β_2, the more peaked the distribution is.

Sometimes $\gamma_2 = \beta_2 - 3$ is taken as a measure of kurtosis.

For a normal distribution, $\gamma_2 = 0$

If $\gamma_2 > 0$, the curve is more peaked (*i.e.*, leptokurtic)

and if $\gamma_2 < 0$, the curve is more flat-topped (*i.e.*, platykurtic).

EXERCISES

1. The following table gives the dist. of plots according to their sizes in a given region. Calculate the quartile deviation. (Size of the form is rounded to the nearest acre).

Farm size (in sq. metres)	*No. of farms*	*Farm size (in sq. metres)*	*No. of farms*
0—40	394	161—200	169
41—80	461	201—240	113
81—120	391	241 and over	148
121—160	334		

Also calcmate the quartile co-efficient of dispersion.

[**Ans.** 50·96; 0·51]

2. The following table gives the frequency dist. of 290 workers of a factory according to their average monthly income in 2000-01.

Income group	*No. of workers*	*Income group*	*No. of workers*
Below 50	1	150—170	22
50—70	16	170—190	15
70—90	39	190—210	15
90—110	58	210—230	9
110—130	60	230 and above	10
130—150	46		

Locate the quartiles and hence calculate co-efficient of dispersion.

[**Ans.** 95·78; 149·24; 0·22]

3. Calculate the S.D. of each of the following sequences of binomial co-efficients :
   (*i*) 1, 5, 10, 5, 1.
   (*ii*) 1, 6, 15, 20, 15, 6, 1.
   (*iii*) 1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1.

[**Ans.** 3·323; 6·958; 90·17]

4. Calculate the second momei following dist. :

x :	0	1	2
f:	1	9	26

5. Calculate the S.D., for the fc emplcyees of a big factory :

Wage (in Rs.)	50—
No. of employees	25(
	25-3
	110

6. Calculate the S.D. of the follc

Marks	*No. o*
More than 0	
More than 10	
More than 20	
More than 30	

7. Calculate the variance of the

Marks	*No. o*
30—35	
35—40	
40—45	
45—50	

8. Calculate the variance of the

Marks	*No. o*
10—14	
14—18	
18—22	
22—26	
26—30	
30—34	

9. Find the standard deviation a

Wages	*No. o*
Up to Rs. 10	
Up to Rs. 20	
Up to Rs. 30	
Up to Rs. 40	1

10. Find out (*a*) median co-effici from the following data :

ve same variability as measured by
e equally flat at the top. The flatness
normal curve. This relative flatness

frequency curve is more peaked or
by β_2. For normal curve, value of

al curve, $\beta_2 > 3$. Such curves are

ped than the normal curve, $\beta_2 < 3$.

distribution is.

tosis.

:)

atykurtic).

ng to their sizes in a given region.
rounded to the nearest acre).

size	No. of farms
netres)	
-200	169
-240	113
nd over	148

[**Ans.** 50·96; 0·51]

workers of a factory according to

group	No. of workers
170	22
190	15
210	15
230	9
d above	10

of dispersion.

[**Ans.** 95·78; 149·24; 0·22]

s of binomial co-efficients :

[**Ans.** 3·323; 6·958; 90·17]

4. Calculate the second moment about the mean and co-efficient of variation of the following dist. :

x :	0	1	2	3	4	5	6	7	8
f :	1	9	26	59	72	52	26	7	1

[**Ans.** 1·98; 35·4%]

5. Calculate the S.D., for the following data relating to the weekly wage dist. of 5000 employees of a big factory :

Wage (in Rs.)	50—55	45—50	40—45	35—40	30—35	25-30	20-25
No. of employees	250	300	400	450	800	1100	1700

[**Ans.** 9·024]

6. Calculate the S.D. of the following dist. :

Marks	*No. of students*	*Marks*	*No. of students*
More than 0	100	More than 40	25
More than 10	90	More than 50	15
More than 20	75	More than 60	5
More than 30	50	More than 70	0

[**Ans.** 15·94]

7. Calculate the variance of the following dist :

Marks	*No. of students*	*Marks*	*No. of students*
30—35	5	50—55	16
35—40	7	55—60	12
40—45	8	60—65	7
45—50	20	65—70	5
			80

[**Ans.** 82·44]

8. Calculate the variance of the following data :

Marks	*No. of students*	*Marks*	*No. of students*
10—14	2	34—38	10
14—18	4	38—42	8
18—22	4	42—46	4
22—26	8	46—50	6
26—30	12	50—54	2
30—34	16	54—58	4

[**Ans.** 110·15]

9. Find the standard deviation and co-efficient of variation from the following data :

Wages	*No. of persons*	*Wages*	*No. of persons*
Up to Rs. 10	12	Up to Rs. 50	157
Up to Rs. 20	30	Up to Rs. 60	202
Up to Rs. 30	65	Up to Rs. 70	222
Up to Rs. 40	107	Up to Rs. 80	230

[**Ans.** 17·26; 42·69%]

10. Find out (*a*) median co-efficient of dispersion and (*b*) mean co-efficient of dispersion from the following data :

Age-group (in years)	No. of men	Age-group (in years)	No. of men
15—20	5	35—40	49
20—25	9	40—45	28
25—30	82	45—50	6
30—35	58	50—55	3

[**Ans.** 0·17; 0·16]

11. Calculate co-efficient of variation of the marks of 40 students given below :

Marks	Students	Marks	Students
80—84	1	50—54	6
75—79	1	45—49	6
70—74	1	40—44	6
65—69	4	35—39	3
60—64	4	30—34	0
55—59	7	25—29	1
			40

[**Ans.** 21·79%]

12. A sample of 5 items is taken from the production of a firm. Length and weight of the 5 items are given below :

Length :	3	4	6	7	10
Weight :	9	11	14	15	16

By comparing the co-efficients of variation of two characters, conclude which of them is more variable.

[**Ans.** Length]

13. During the first 10 weeks of a session the marks of two students X and Y taking the course were :

X :	58	59	60	54	65	66	52	75	69	52
Y :	56	87	89	78	71	73	84	65	65	46

Which of the two is more consistent ?

[**Ans.** X]

14. The scores of two golfers for 24 rounds each are :

A :	74	75	78	78	72	77	79	78	81	76	72	72	77
	74	70	78	79	80	81	74	80	75	71	73		
B :	86	84	80	88	89	85	86	82	82	79	86	80	82
	76	86	89	87	83	80	88	86	81	84	87		

Which may be considered to be more consistent?

[**Ans.** B]

15. Goals scored by two teams A and B in a football season were as follows :

No. of Goals scored in a match :	0	1	2	3	4
No. of matches by A :	54	18	16	10	8
No. of matches by B :	34	18	12	10	6

Which team is more consistent and why?

[**Ans.** B]

16. The following table gives the dist. of house-holds according to size in two cities A and B.

Size of house-hold	City A	City B
1	24	14
2	10	10
3	12	12
4	15	13
5	13	14

6
7
8

Derive a measure to study the

17. The index numbers of prices
follows :

Month In
(

Jan.
Feb.
March
April
May
June
July
August
September
October
November
December

Which of the two shares do y

18. Calculate Karl Pearson's co-

Marks No. o
Above 0
Above 10
Above 20
Above 30

19. The table below gives ages of

Age (in months)
15—16
17—18
19—20
21—22
23—24
25—26
27—28
29—30
31—32
33—34
35—36
37—38

roup ars)	No. of men
40	49
45	28
50	6
55	3

[**Ans.** 0·17; 0·16]

students given below :

ks	*Students*
54	6
49	6
44	6
39	3
34	0
29	1
	40

[**Ans.** 21·79%]

irm. Length and weight of the 5

7	10
15	16

acters, conclude which of them

[**Ans.** Length]

vo students X and Y taking the

52	75	69	52
84	65	65	46

[**Ans.** X]

81	76	72	72	77
75	71	73		
82	79	86	80	82
81	84	87		

[**Ans.** B]

ı were as follows :

2	3	4
16	10	8
12	10	6

[**Ans.** B]

ording to size in two cities A

B

6	10	11
7	6	10
8	10	16
	100	100

Derive a measure to study the variability of the dist.

[**Ans.** A : 59·66%; B : 51·47%]

17. The index numbers of prices of cotton and jute shares in a particular year were as follows :

Month	*Index no. of prices of cotton shares*	*Index no. of prices of jute shares*
Jan.	188	131
Feb.	178	130
March	173	130
April	164	129
May	172	129
June	183	120
July	184	127
August	185	127
September	211	130
October	217	137
November	232	140
December	240	142

Which of the two shares do you consider to be more variable in price ?

[**Ans.** Cotton shares]

18. Calculate Karl Pearson's co-efficient of skewness from the following data :

Marks	*No. of students*	*Marks*	*No. of students*
Above 0	1500	Above 40	780
Above 10	1400	Above 50	700
Above 20	1000	Above 60	300
Above 30	780	Above 70	140
		Above 80	0

[**Ans.** 0·995]

19. The table below gives ages of children in two nursery schools. Compare their variability :

Age (in months)	*No. of children school A*	*No. of children school B*
15—16	—	1
17—18	1	2
19—20	2	2
21—22	5	5
23—24	7	10
25—26	9	7
27—28	8	6
29—30	5	3
31—32	3	3
33—34	1	1
35—36	1	—
37—38	—	2
	42	42

(**Ans.** Ages of children in school B are more variable,

20. Find the co-efficient of variation and Pearson's co-efficient of skewness from the following data :

Year	*Price Index No. of wheat*	*Year*	*Price Index No. of wheat*
1990	83	1995	126
1991	87	1996	130
1992	93	1997	118
1993	100	1998	106
1994	124	1999	104

(**Ans.** 14·85; 0·396)

21. Find Pearson's measure of skewness for the data given below :

Weight (in lbs)	*No. of persons*	*Weight (in lbs)*	*No. of persons*
70— 79·99	12	110—119·99	50
80— 89·99	18	120—129·99	45
90— 99·99	35	130—139·99	20
100—109·99	42	140—149·99	8

(**Ans.** –5·719)

22. Find the co-efficient of dispersion and Pearson's co-efficient of skewness for the following data :

Wage (in Rs.)	*No. of persons*	*Wage (in Rs.)*	*No. of persons*
70— 80	12	110—120	50
80— 90	18	120—130	45
90—100	35	130—140	20
100—110	42	140—150	8
			230

(**Ans.** 0·16; –0·33)

23. For data in Ex. 2 locate median, quartiles and hence co-efficient of skewness.

(**Ans.** 0·08)

24. Calculate the Pearson's co-efficient of skewness for the following dist. of weights of boys :

Weight (in kgs.)	*Frequency*	*Weight (in kgs.)*	*Frequency*
20·5 – 23·5	17	29·5 – 32·5	194
23·5 – 26·5	193	32·5 – 35·5	27
26·5 – 29·5	399	35·5 – 38·5	10

(**Ans.** 0·07)

25. The first three moments about the origin are

$$\mu_1' = \frac{1}{2}(n+1), \quad \mu_2' = \frac{1}{6}(n+1)(2n+1)$$

$$\mu_3' = \frac{1}{4}n(n+1)^2$$

Examine the skewness of the data.

26. Let x be a random variable with mean μ and variance σ^2. Show that E $\{(x-b)^2\}$ as f^n of b, is minimized when $b = \mu$.

27. Show that the co-efficient of skewness ranges from –1 to 1

(**Hint :** use $Q_1 < Q_2 < Q_3$).

□□

Theory and As

3.1. Introduction

Attribute means quality or prope
the several attributes. An object or ir
and the class of individual possessin
.... are used to denote the absence of
the attribute **honesty,** *α represents* **di**
by grouping the letters representin
blindness, AB represents combinatio
Any letter or combination of lett
members of a class are specified are
Class-frequencies. *The numb*
frequency. It is denoted by enclosing
denotes the number of objects or inc
Order of Classes and Class-fr
a class of order r and its frequency a
and second order and (A), (AB) are
The total frequency is denoted i
Ultimate Class-frequencies. *T*
frequencies.
Positive and Negative Attribu
positive attributes and those denot
symbol includes only capital letters
and if only Greek letters, a negative
α β is negative class.
Symbol. *A symbol 'A.N.' is use*
A.N = (
which is the symbolic way of saying
(A) is obtained.
Condition of Consistence. *Th*
of a set of independent class-freque
Ex. 3-1. *Show that if there are*
Sol. There is only one class of
classes of order one are
$A_1, A_2, \ldots A_n, \alpha_1, \alpha_2, \ldots \alpha_n$ (c
number.
To find classes of order 2, cons
the classes
A_1A

·efficient of skewness from the

r	Price Index No. of wheat
5	126
6	130
7	118
8	106
9	104

(Ans. 14·85; 0·396)

below :

ht s)	No. of persons
9·99	50
9·99	45
9·99	20
9·99	8

(Ans. –5·719)

-efficient of skewness for the

)	No. of persons
20	50
30	45
40	20
50	8
	230

(Ans. 0·16; –0·33)

-efficient of skewness.

(Ans. 0·08)

following dist. of weights of

)	Frequency
:·5	194
·5	27
·5	10

(Ans. 0·07)

$\iota+1)(2n+1)$

^{2}. Show that E $\{(x-b)^2\}$ as

) 1

(Hint : use $Q_1 < Q_2 < Q_3$).

□□

3

Theory and Association of Attributes

3.1. Introduction

Attribute means quality or property. The capital letters A, B, C, ... are used to denote the several attributes. An object or individual possessing the attribute A is termed simply A and the class of individual possessing A is termed as **the class A.** *The Greek letters* α, β, γ *.... are used to denote the absence of attributes A, B, C, ... respectively e.g., if A represents the attribute* **honesty,** α *represents* **dishonesty**. *The combination of attributes is represented by grouping the letters representing the attributes, e.g., if A represents deafness and B blindness, AB represents combination deafness and blindness.*

Any letter or combination of letters like A, AB by means of which the characters of the members of a class are specified are called **class-symbol.**

Class-frequencies. *The number of observations in any class is called the class-frequency. It is denoted by enclosing the corresponding class-symbol in brackets. e.g., (A) denotes the number of objects or individuals possessing the attribute A.*

Order of Classes and Class-frequencies. *A class specified by r attributes is known as a class of order r and its frequency as a frequency of rth order, e.g., A, AB are classes of first and second order and (A), (AB) are frequencies of order one and two respectively.*

The total frequency is denoted by N.

Ultimate Class-frequencies. *The frequency of highest order are called ultimate class-frequencies.*

Positive and Negative Attributes. *The attributes denoted by capitals are termed as positive attributes and those denoted by Greek letters as negative attributes. If a class-symbol includes only capital letters, the corresponding class is termed as a positive class and if only Greek letters, a negative class e.g., the class AB is positive class and the class* $\alpha\beta$ *is negative class.*

Symbol. *A symbol 'A.N.' is used for the dichotomising N according to A and is written*

$$A.N = (A)$$

which is the symbolic way of saying that if N is dichotomised according to A, class-frequency (A) is obtained.

Condition of Consistence. *The necessary and sufficient condition for the consistency of a set of independent class-frequencies is that no ultimate class-frequency is negative.*

Ex. 3-1. *Show that if there are n attributes, the number of distinct classes is* 3^n.

Sol. There is only one class of order zero, If $A_1, A_2, \ldots A_n$ be n attributes the possible classes of order one are

$A_1, A_2, \ldots A_n$, $\alpha_1, \alpha_2, \ldots \alpha_n$ (α's denoting the absence of A's) which are $2n = {}^nc_1 . 2$ in number.

To find classes of order 2, consider two attributes A_1 and A_2. These two attributes give the classes

$$A_1A_2,\ A_1\alpha_2,\ \alpha_1A_2 \text{ and } \alpha_1\alpha_2$$

of order two. These are $2^2 = 4$ in number. Since out of n two attributes can be chosen in ${}^nc_2$ ways, total number of classes of order two

$$= {}^nc_2 \,.\, 2^2$$

Similarly the number of classes of order 3

$$= {}^nc_3 \,.\, 2^3$$

and in general the number of classes of order r

$$= {}^nc_r \,.\, 2^r$$

$\therefore$ Total number of classes

$$= 1 + {}^nc_1 \,.\, 2 + {}^nc_2 \,.\, 2^2 + \ldots + {}^nc_r \,.\, 2^r + \ldots + {}^nc_n \,.\, 2^n$$
$$= (1+2)^n = 3^n.$$

Ex. 3-2. *A number of school-children were examined for the presence or absence of certain defects of which three chief descriptions were noted : A, development defect; B, nerve signs; C, low nutrition. Given the following ultimate frequencies, find the frequencies of the classes defined by the presence of the defects i.e., those involving the Roman letters. A, B, C but not the Greek letters α, β, γ including the whole number of observations N.*

(ABC)	57	(αBC)	78
$(AB\gamma)$	281	$(\alpha B\gamma)$	670
$(A\beta C)$	86	$(\alpha\beta C)$	65
$(A\beta\gamma)$	453	$(\alpha\beta\gamma)$	8310

Sol. N = Total no. of observations

= Sum of frequencies given above = 10,000

$$
\begin{aligned}
\text{Now} \qquad (A) &= (AB) + (A\beta)\\
(AB) &= (ABC) + (AB\gamma)\\
(A\beta) &= (A\beta C) + (A\beta\gamma)\\
\therefore \qquad (A) &= (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma)\\
&= 57 + 281 + 86 + 453 = 877\\
\text{Similarly} \qquad (B) &= (ABC) + (AB\gamma) + (\alpha BC) + (\alpha B\gamma)\\
&= 57 + 281 + 78 + 670 = 1{,}086\\
(C) &= (ABC) + (A\beta C) + (\alpha BC) + (\alpha\beta C)\\
&= 57 + 86 + 78 + 65 = 286\\
(AB) &= (ABC) + (AB\gamma) = 57 + 281 = 338\\
(BC) &= (ABC) + (\alpha BC) = 57 + 78 = 135\\
(AC) &= (ABC) + (A\beta C) = 57 + 86 = 143\\
\text{Also} \qquad (ABC) &= 57.
\end{aligned}
$$

Ex. 3-3. *From the frequencies given below find all the class-frequencies.*

N = 10,000,	$(A) = 877$,	$(B) = 1086$,	$(C) = 286$
$(AB) = 338$	$(AC) = 143$,	$(BC) = 135$,	$(ABC) = 57$

$$
\begin{aligned}
\textbf{Sol.}\ \text{Now} \qquad (AB) &= (ABC) + (AB\gamma)\\
\therefore \qquad (AB\gamma) &= (AB) - (ABC) = 338 - 57 = 281\\
\text{Similarly} \qquad (A\beta C) &= -(ABC) + (AC) = 143 - 57 = 86\\
(\alpha BC) &= (BC) - (ABC) = 135 - 57 = 78\\
\text{Now} \qquad (A\beta\gamma) &= (A\beta) - (A\beta C)\\
&= (A) - (AB) - (A\beta C)\\
&= 877 - 338 - 86 = 453\\
(\alpha\beta C) &= (\beta C) - (A\beta C) = (C) - (BC) - (A\beta C) = 286 - 135 - 86 = 65\\
(\alpha\beta\gamma) &= (\alpha\beta) - (\alpha\beta C)\\
&= (\alpha) - (\alpha B) - (\alpha\beta C)\\
&= \{N - (A)\} - \{(B) - (AB)\} - (\alpha\beta C)
\end{aligned}
$$

=

$(\alpha B\gamma)$ =

=

Thus all the ultimate freque

obtained as in last example.

Ex. 3.4. *Show that* $A + \alpha$ =

$\alpha . N = (\alpha)$ *and* $1 . N = N$ *Deduce*

$(\alpha\beta\gamma)$ =

Sol. By def. $A . N = (A)$ ar

$\therefore$ $A . N + \alpha . N$ =

or $(A + \alpha) . N$ =

$\therefore$ $A + \alpha$ =

$\therefore$ α =

Similarly β =

$\therefore$ $(\alpha\beta\gamma)$ =

Ex. 3-5. *Given that*

(A)

Show that $(AB) = (\alpha\beta)$, $(A\beta$

Sol. (AB)

$(A\beta)$

Ex. 3-6. *Given that* (A)

and also that (ABC)

show that $2(ABC)$

Sol. (ABC)

$\therefore$ $2(ABC)$

Ex. 3-7. *Measurements are*

measurements of the husbands

measurements, in 700 cases fo

many cases will both measurem

Sol. Let A and B denote the

respectively. Then

N

$\therefore$ $(\alpha\beta)$

ı two attributes can be chosen in $^nc_2$

$+ \ldots + {}^nc_r \,.\, 2^r + \ldots + {}^nc_n \,.\, 2^n$

ned for the presence or absence of
noted : A, development defect; B,
ıte frequencies, find the frequencies
those involving the Roman letters.
hole number of observations N.

78
670
65
8310

,000

$) + (A\beta\gamma)$
$= 877$
$) + (\alpha B\gamma)$
$= 1{,}086$
$C) + (\alpha\beta C)$
86
$281 = 338$
$- 78 = 135$
$- 86 = 143$

he class-frequencies.

6, $(C) = 286$
5, $(ABC) = 57$

$57 = 281$
$- 57 = 86$
$57 = 78$

$C) - (A\beta C) = 286 - 135 - 86 = 65$

$)\} - (\alpha\beta C)$

$$\begin{aligned}
&= 10{,}000 - 877 - 1086 + 338 - 65 = 8310\\
(\alpha B\gamma) &= (\alpha B) - (\alpha BC)\\
&= (B) - (AB) - (\alpha BC) = 1086 - 338 - 78 = 670.
\end{aligned}$$

Thus all the ultimate frequencies are known. From these remaining frequencies can be obtained as in last example.

Ex. 3.4. *Show that $A + \alpha = 1$ where A, α and 1 are operators defined by $A.N. = (A)$, $\alpha.N = (\alpha)$ and $1 \,.\, N = N$ Deduce that*

$$(\alpha\beta\gamma) = N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC)$$

Sol. By def. $A \,.\, N = (A)$ and $\alpha \,.\, N = (\alpha)$

$$\begin{aligned}
\therefore \quad A \,.\, N + \alpha \,.\, N &= (A) + (\alpha) = N\\
\text{or} \quad (A + \alpha) \,.\, N &= 1 \,.\, N\\
\therefore \quad A + \alpha &= 1\\
\therefore \quad \alpha &= 1 - A.\\
\text{Similarly} \quad \beta &= 1 - B \text{ and } \gamma = 1 - C\\
\therefore \quad (\alpha\beta\gamma) &= \alpha\beta\gamma \,.\, N = (1 - A)\,(1 - B)\,(1 - C) \,.\, N\\
&= (1 - A - B - C + AB + AC + BC - ABC) \,.\, N\\
&= 1.N - A.N - B.N - C.N + AB.N + AC.N + BC.N - ABC.N\\
&= N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC).
\end{aligned}$$

Ex. 3-5. *Given that*

$$(A) = (\alpha) = (B) = (\beta) = \frac{N}{2}$$

Show that $(AB) = (\alpha\beta)$, $(A\beta) = (\alpha B)$.

Sol.

$$\begin{aligned}
(AB) &= AB.N = (1 - \alpha)\,(1 - \beta).N\\
&= \{1 - \alpha - \beta + \alpha\beta\}.N\\
&= N - (\alpha) - (\beta) + (\alpha\beta)\\
&= N - \frac{N}{2} - \frac{N}{2} + (\alpha\beta) = (\alpha\beta)\\
(A\beta) &= A\beta.\text{N} = \{1 - \alpha\}\ \{1 - B).N\\
&= \{1 - \alpha - B + \alpha B\}.N\\
&= N - (\alpha) - (B) + (\alpha B) = (\alpha B).
\end{aligned}$$

Ex. 3-6. *Given that* $(A) = (\alpha) = (B) = (\beta) = (C) = (\gamma) = \dfrac{N}{2}$

and also that $(ABC) = (\alpha\beta\gamma)$

show that $2(ABC) = (AB) + (AC) + (BC) - \dfrac{N}{2}$

Sol.

$$\begin{aligned}
(ABC) &= (\alpha\beta\gamma) = N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC)\\
\therefore \quad 2(ABC) &= (AB) + (AC) + (BC) - \frac{N}{2}.
\end{aligned}$$

Ex. 3-7. *Measurements are made on a thousand husbands and a thousand wives. If the measurements of the husbands exceed the measurements of the wives in 800 cases for one measurements, in 700 cases for another and in 660 cases for both measurements, in how many cases will both measurements on the wife exceed the measurements on the husband ?*

Sol. Let A and B denote the husbands exceeding wives in first and second measurements respectively. Then

$$\begin{aligned}
N &= 1000,\ (A) = 800,\ (B) = 700 \text{ and } (AB) = 660\\
\therefore \quad (\alpha\beta) &= \alpha\beta.N = \{1 - A\}\ \{1 - B\}.N\\
&= \{1 - A - B + AB\}.N = N - (A) - (B) + (AB)\\
&= 1000 - 800 - 700 + 660 = 160.
\end{aligned}$$

Ex. 3-8. 100 *children took three examinations,* 40 *passed the first,* 39 *passed the second and* 48 *passed the third.* 10 *passed all three,* 21 *failed all three,* 9 *passed the first two and failed the third,* 19 *failed the first two and passed the third. Find how many children passed at least two examinations.*

Show that for the question asked certain of the given frequencies are not necessary. Which are they ?

Sol. Let A, B, C denote passing first, second and third examinations respectively. Then

$$N = 100, (A) = 40, (B) = 39, (C) = 48, (ABC) = 10$$

$$(\alpha\beta\gamma) = 21,\ (AB\gamma) = 9 \text{ and } (\alpha\beta C) = 19$$

Now
$$\begin{aligned}(\alpha BC) &= \alpha BC.N = (1-A)\,BCN = (BC - ABC).N\\ &= (BC) - (ABC)\end{aligned}$$

and
$$\begin{aligned}(A\beta C) &= A\beta C.N = (1-\alpha)\,\beta C.N = \{\beta C - \alpha\beta C\}.N\\ &= (\beta C) - (\alpha\beta C)\end{aligned}$$

$\therefore$ *No. of children who passed at least two examinations*

$$\begin{aligned}&= (\alpha BC) + (A\beta C) + (AB\gamma) + (ABC)\\ &= (BC) - (ABC) + (\beta C) - (\alpha\beta C) + (AB\gamma) + (ABC)\\ &= (C) - (\alpha\beta C) + (AB\gamma) = 48 - 19 + 9 = 38.\end{aligned}$$

Evidently three frequencies have been used and hence others are not necessary.

Ex. 3-9. *Show that if A occurs in a larger proportion of the cases where B is than where B is not, then B will occur in a larger proportion of the cases where A is than where A is not.*

Sol. It is given that

$$\frac{(AB)}{(B)} > \frac{(A\beta)}{(\beta)}$$

and it is to be shown that

$$\frac{(AB)}{(A)} > \frac{(\alpha B)}{(\alpha)}$$

From given
$$\frac{(\beta)}{(B)} > \frac{(A\beta)}{(AB)}$$

$\therefore$
$$1 + \frac{(\beta)}{(B)} > 1 + \frac{(A\beta)}{(AB)}$$

or
$$\frac{(B)+(\beta)}{(B)} > \frac{(AB)+(A\beta)}{(AB)}$$

or
$$\frac{N}{(B)} > \frac{(A)}{(AB)}$$

or
$$\frac{N}{(A)} > \frac{(B)}{(AB)}$$

or
$$\frac{(A)+(\alpha)}{(A)} > \frac{(AB)+(\alpha B)}{(AB)}$$

or
$$1 + \frac{(\alpha)}{(A)} > 1 + \frac{(\alpha B)}{(AB)}$$

or
$$\frac{(\alpha)}{(A)} > \frac{(\alpha B)}{(AB)}$$

or
$$\frac{(AB)}{(A)} > \frac{(\alpha B)}{(\alpha)}$$

Ex. 3-10. *At a competitive*
outnumbered girls by 96. Those qu
to qualify by 310. The number of sc
the Art graduate girls there were 2
were only 135 Art graduates and 33
numbered 18. Find out.
(*i*) *the number of boys who*
(*ii*) *the total number of scie*
(*iii*) *the number of science g*
Sol. Let A, B and C denote th
science candidate. Then
$N = 600,\ (A) - (\alpha) = 96,\ (B)$ -
$(\gamma) = 135,\ (\beta\gamma) = 33$ and $(A\beta)$
Since $N = (A) + (\alpha) = 600 = ($
$(A) = 348,\ (\alpha) = 252,\ (B) = 4$
$\therefore$ (*i*) $(AB) = (A) -$
(*ii*) $(AC) = (ABC$
$= (ABC$
$= 300 +$
(*iii*) $(\alpha BC) = (BC) -$
$= N - (\gamma$
$= 600 -$

Ex. 3-11. *In a free vote in the*
members representing English cons
25 Opposition members representi
Government majority among those
Scottish constituencies. 18 Gover
members voted in favour of the mo
voting according to the nationality
Sol. Let A, B, C denote the a
motion and being members of Eng
Then $N = 600, (ABC) = 300, ($
and $(B) - (\beta) = 310.$
It is required to find all ultima
Now $N =$
$\therefore$ $(A) =$
$\therefore$ $(C) =$
$(AB) =$
$=$
$(BC) =$
$(\alpha\beta\gamma) =$
$\therefore$ $(AC) =$
$=$
$=$
Now $(AB\gamma) =$
$(A\beta C) =$
$(\alpha BC) =$
$(A\beta\gamma) =$
$(\alpha\beta C) =$
$=$

the first, 39 *passed the second*
ee, 9 *passed the first two and*
nd how many children passed

equencies are not necessary.

aminations respectively. Then
$) = 48, (ABC) = 10$
$= 19$
$3C - ABC).N$

$\beta C - \alpha\beta C\}.N$

ons
(ABC)
$\beta C) + (AB\gamma) + (ABC)$
$- 19 + 9 = 38.$
thers are not necessary.
e cases where B is than where
here A is than where A is not.

Ex. 3-10. *At a competitive examination at which* 600 *graduates appeared, boys outnumbered girls by* 96. *Those qualifying for interview exceeded in number those failing to qualify by* 310. *The number of science graduate boys interviewed was* 300 *while among the Art graduate girls there were* 25 *who failed to qualify for interview. All together there were only* 135 *Art graduates and* 33 *among them failed to qualify. Boys who failed to qualify numbered* 18. *Find out.*

(*i*) *the number of boys who qualified for interview.*
(*ii*) *the total number of science graduate boys appearing.*
(*iii*) *the number of science graduate girls who qualified.*

Sol. Let A, B and C denote the attributes of being a boy, qualified for interview and science candidate. Then

$N = 600,\ (A) - (\alpha) = 96,\ (B) - (\beta) = 310, (ABC) = 300, (\alpha\beta\gamma) = 25,$
$(\gamma) = 135, (\beta\gamma) = 33$ and $(A\beta) = 18.$

Since $N = (A) + (\alpha) = 600 = (B) + (\beta) = (C) + (\gamma)$
$(A) = 348,\ (\alpha) = 252,\ (B) = 455,\ (\beta) = 145, (C) = 465;$

$$\begin{aligned}
\therefore \quad (i) \quad (AB) &= (A) - (A\beta) = 348 - 18 = 330 \\
(ii) \quad (AC) &= (ABC) + (A\beta C) = (ABC) + (A\beta) - (A\beta\gamma) \\
&= (ABC) + (A\beta) - (\beta\gamma) + (\alpha\beta\gamma) \\
&= 300 + 18 - 33 + 25 = 310 \\
(iii) \quad (\alpha BC) &= (BC) - (ABC) = \{1 - \beta\} \cdot (1 - \gamma).\ N - (ABC) \\
&= N - (\gamma) - (\beta) + (\beta\gamma) - (ABC) \\
&= 600 - 135 - 145 + 33 - 300 = 53.
\end{aligned}$$

Ex. 3-11. *In a free vote in the house of commons,* 600 *members voted.* 300 *Government members representing English constituencies (including welsh) voted in favour of the motion.* 25 *Opposition members representing Scottish constituencies voted against the motion. The Government majority among those who voted was* 96. 135 *of the members voting represented Scottish constituencies.* 18 *Government members voted against the motion.* 102 *Scottish members voted in favour of the motion. The motion was carried by* 310 *votes. Analyse the voting according to the nationality of the constituencies and party.*

Sol. Let A, B, C denote the attributes of being Government members, voting for the motion and being members of English constituencies respectively.

Then $N = 600, (ABC) = 300, (\alpha\beta\gamma) = 25,\ (A) - (\alpha) = 96, (\gamma) = 135, (A\beta) = 18, (B\gamma) = 102$ and $(B) - (\beta) = 310.$

It is required to find all ultimate frequencies.

$$\begin{aligned}
\text{Now} \qquad N &= (A) + (\alpha) = 600 = (B) + (\beta) \\
\therefore \qquad (A) &= 348, (\alpha) = 252, (B) = 455 \text{ and } (\beta) = 145. \\
\therefore \qquad (C) &= N - (\gamma) = 465. \\
(AB) &= AB.N = A(1 - \beta).N = (A) - (A\beta) \\
&= 348 - 18 = 330 \\
(BC) &= (B) - (B\gamma) = 455 - 102 = 353 \\
(\alpha\beta\gamma) &= N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC) \\
\therefore \qquad (AC) &= (\alpha\beta\gamma) + (ABC) + (A) + (B) + (C) - (BC) - (AB) - N \\
&= 25 + 300 + 348 + 455 + 465 - 353 - 330 - 600 \\
&= 310. \\
\text{Now} \qquad (AB\gamma) &= (AB) - (ABC) = 330 - 300 = 30 \\
(A\beta C) &= (AC) - (ABC) = 310 - 300 = 10 \\
(\alpha BC) &= (BC) - (ABC) = 353 - 300 = 53 \\
(A\beta\gamma) &= (A\beta) - (A\beta C) = 18 - 10 = 8 \\
(\alpha\beta C) &= \{1 - A\} \cdot \{1 - B\}\ C.N = \{C - AC - BC + ABC\}.N \\
&= (C) - (AC) - (BC) + (ABC)
\end{aligned}$$

$$= 465 - 310 - 353 + 300 = 102$$

$$(\alpha B\gamma) = (B) - (AB) - (BC) + (ABC)$$
$$= 455 - 330 - 353 + 300 = 72.$$

Ex. 3-12. *Prove that in case of two attributes A and B, the conditions of consistency are* (*i*) $(AB) \geq 0$ (*ii*) $(AB) \leq (A)$ (*iii*) $(AB) \leq (B)$ and (*iv*) $(AB) \geq (A) + (B) - N$.

Sol. The necessary and sufficient condition for the consistency is that no ultimate frequency is negative *i.e.*, $(AB) \geq 0$, $(A\beta) \geq 0$, $(\alpha B) \geq 0$ and $(\alpha\beta) \geq 0$.

Now
$$(A\beta) = \{A\beta\}.N = A\{1 - B\}.N = A.N - \{AB\} \cdot N$$
$$= (A) - (AB)$$

$\therefore$ $(A\beta) \geq 0$ implies $(A) \geq (AB)$

Similarly $(\alpha B) \geq 0$ implies $(B) \geq (AB)$

Now
$$(\alpha\beta) = \{\alpha\beta\}.N = \{(1 - A)(1 - B)\}.N$$
$$= \{1 - A - B + AB\}.N = N - (A) - (B) + (AB)$$

$\therefore$ $(\alpha\beta) \geq 0$ implies $(AB) \geq (A) + (B) - N$.

Ex. 3-13. *Prove that in case of three attributes A, B and C, the conditions of consistency are*

(*i*) $(ABC) \geq 0$ (*ii*) $(ABC) \geq (AB) + (AC) - (A)$
(*iii*) $(ABC) \geq (AB) + (BC) - (B)$ (*iv*) $(ABC) \geq (AC) + (BC) - (C)$
(*v*) $(ABC) \leq (AB)$ (*vi*) $(ABC) \leq (AC)$
(*vii*) $(ABC) \leq (BC)$
and (*viii*) $(ABC) \leq (AB) + (AC) + (BC) - (A) - (B) - (C) + N$

Hence deduce

$(AB) + (AC) + (BC) \geq (A) + (B) + (C) - N$
$(AB) + (AC) - (BC) \leq (A)$
$(AB) - (AC) + (BC) \leq (B)$
$-(AB) + (AC) + (BC) \leq (C)$

Sol. For consistency it is necessary and sufficient that all ultimate frequencies are non-negative *i.e.*,

(*i*) $(ABC) \geq 0$
(*ii*) $(A\beta\gamma) \geq 0$ *i.e.*, $\{A\beta\gamma\}.N \geq 0$
i.e., $\{A(1 - B)(1 - C)\}.N \geq 0$
i.e., $\{A - AB - AC + ABC\}.N \geq 0$
i.e., $(A) - (AB) - (AC) + (ABC) \geq 0.$
i.e., $(ABC) \geq (AB) + (AC) - (A)$

Similarly $(\alpha B\gamma) \geq 0$ and $(\alpha\beta C) \geq 0$ implies (*iii*) and (*iv*).

(*v*) $(AB\gamma) \geq 0$
i.e., $\{AB\gamma\}.N \geq 0$
i.e., $AB(1 - C).N \geq 0$
i.e., $\{AB - ABC\}.N \geq 0$
i.e., $(AB) \geq (ABC)$.

Similarly (*vi*) and (*vii*) follow from $(A\beta C) \geq 0$ and $(\alpha BC) \geq 0$.

(*viii*) $(\alpha\beta\gamma) \geq 0$
i.e., $\{1 - A\} \cdot \{1 - B\} \cdot \{1 - C\}.N \geq 0$
i.e., $\{1 - A - B - C + AB + AC + BC - ABC\}.N \geq 0$
i.e., $N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC) \geq 0$
i.e., $(ABC) \leq (AB) + (AC) + (BC) - (A) - (B) - (C) + N.$

From (*i*) and (*viii*)
$$(AB) + (AC) + (BC) - (A) - (B) - (C) + N \geq 0$$
i.e., $(AB) + (AC) + (BC) \geq (A) + (B) + (C) - N$

From (*ii*) and (*vii*)
$(AB) + (AC) - (A) \leq (BC$
i.e., $(AB) + (AC) - (BC) \leq (A$
Similarly from (*iii*) and (*vi*), (*i*
$(AB) + (BC) - (AC) \leq (B$
and $(AC) + (BC) - (AB) \leq (C$

Ex. 3-14. *If a report gives the*
there must be a misprint or mistake
$N = 1000$, $(A) = 510$, $(B) = 4$
Sol. Now $(AB) + (AC) + (B$
and $(A) + (B) + (C) -$
$\therefore$ $(AB) + (AC) + (BC) < (A$
$\therefore$ Data is not consistent and l

Ex. 3-15. *If in an urban distr*
years of age were returned as "occ
widowed, what is the lowest propo
have been occupied ?
Sol. Let A and B denote the att
respectively. Then
$N = 1000$, $(A) = 817$ and $(B) =$
$\therefore$ $(AB) \geq (A) + (B) - N =$
$\therefore$ Lowest proportion per tho
occupied
=

Ex. 3-16. *A market investigat*
811 *liked chocolates*, 752 *liked toff*
toffee, 356 *liked chocolates and b*
liked all three. Show that this info
Sol. Let A, B and C denote t
boiled sweets respectively. Then
$N = 1000$, $(A) = 811$, $(B$
and $(ABC) = 297$.
Now $(\alpha\beta\gamma) =$
=
=
$\therefore$ Information is incorrect

Ex. 3-17. *In a war between*
White; there are more armed Whit
ammunition than unarmed White
Reds without ammunition than un
Sol. Let A, B and C denote t
ammunition. Then
$(A) < (\alpha$
and it is to be shown that

Now $(\alpha\beta\gamma) =$

From (*ii*) and (*vii*)

$$(AB) + (AC) - (A) \le (BC)$$

i.e., $(AB) + (AC) - (BC) \le (A)$.

Similarly from (*iii*) and (*vi*), (*iv*) and (*v*)

$$(AB) + (BC) - (AC) \le (B)$$

and $(AC) + (BC) - (AB) \le (C)$.

Ex. 3-14. *If a report gives the following frequencies as actually observed, show that there must be a misprint or mistake of some sort :*

$N = 1000,\ (A) = 510,\ (B) = 490,\ (C) = 427,\ (AB) = 189 \qquad (AC) = 140,\ (BC) = 85$

Sol. Now $(AB) + (AC) + (BC) = 189 + 140 + 85 = 414$

and $(A) + (B) + (C) - N = 510 + 490 + 427 - 1000 = 427$

$\therefore \quad (AB) + (AC) + (BC) < (A) + (B) + (C) - N$

$\therefore$ Data is not consistent and hence there must be a misprint or mistake of some sort.

Ex. 3-15. *If in an urban district* 817 *per thousand of the women between* 20 *and* 25 *years of age were returned as "occupied" at a census and* 263 *per thousand as married or widowed, what is the lowest proportion per thousand of the married or widowed that must have been occupied ?*

Sol. Let A and B denote the attributes of being occupied and being married or widowed respectively. Then

$N = 1000,\ (A) = 817$ and $(B) = 263$

$\therefore \quad (AB) \ge (A) + (B) - N = 817 + 263 - 1000 = 80$

$\therefore$ Lowest proportion per thousand of the married or widowed that must have been occupied

$$= \frac{80}{263} \times 1000 = 304.$$

Ex. 3-16. *A market investigator returns the following data. Of* 1000 *people consulted,* 811 *liked chocolates,* 752 *liked toffee and* 418 *liked boiled sweets;* 570 *liked* chocolates and toffee, 356 *liked chocolates and boiled sweets and* 348 *liked toffee and boiled sweets;* 297 *liked all three. Show that this information as it stands must be incorrect.*

Sol. Let A, B and C denote the attributes of having liking for chocolates, toffee and boiled sweets respectively. Then

$N = 1000,\ (A) = 811,\ (B) = 752,\ (C) = 418,\ (AB) = 570,\ (AC) = 356,\ (BC) = 348$ and $(ABC) = 297$.

Now
$$\begin{aligned}(\alpha\beta\gamma) &= N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC)\\ &= 1000 - 811 - 752 - 418 + 570 + 356 + 348 - 297\\ &= -4 < 0\end{aligned}$$

$\therefore$ Information is incorrect.

Ex. 3-17. *In a war between White and Red forces there are more Red soldiers than White; there are more armed Whites than unarmed Reds : there are fewer armed Reds with ammunition than unarmed Whites without ammunition. Show that there are more armed Reds without ammunition than unarmed Whites with ammunition.*

Sol. Let A, B and C denote the attributes of being. White, armed and possessed with ammunition. Then

$$(A) < (\alpha),\ (AB) > (\alpha\beta),\ (\alpha BC) < (A\beta\gamma)$$

and it is to be shown that

$$(\alpha B\gamma) > (A\beta C)$$

Now
$$\begin{aligned}(\alpha\beta\gamma) &= (\alpha B) - (\alpha BC)\\ &> (\alpha B) - (A\beta\gamma) = (\alpha) - (\alpha\beta) - (A\beta\gamma)\\ &> (A) - (AB) - (A\beta\gamma) = (A\beta) - (A\beta\gamma) = (A\beta C)\end{aligned}$$

Ex. 3-18. *If, in a series of houses actually invaded by smallpox 70% of the inhabitants are attacked and 85% have been vaccinated, what is the lowest percentage of the vaccinated that must have been attacked?*

Sol. Let A and B denote the attributes of being attacked and vaccinated. Then

$$N = 100,\ (A) = 70 \text{ and } (B) = 85$$

Now $(AB) > (A) + (B) - N = 70 + 85 - 100 = 55$

$\therefore$ Lowest percentage of the vaccinated that have been attacked

$$= \frac{55}{85} \times 100 = 65\% \text{ (approx.)}$$

Ex. 3-19. *If all A's are B's and all B's are C's, show that all A's are C's.*

Sol. It is given that $(AB) = (A)$, $(BC) = (B)$

and it is to be shown that

$$(AC) = (A)$$

From Ex. 3-13, $(AB) - (AC) + (BC) \le (B)$

$\therefore \quad (A) - (AC) + (B) \le (B)$

$\therefore \quad (A) \le (AC)$

But $\quad (AC) \le (A)$

$\therefore \quad (AC) = (A).$

Ex. 3-20. *If all A's are B's and no B's are C's, show that no A's are C's.*

Sol. It is given that $(AB) = (A)$ and $(BC) = 0$

and it is to be proved that $(AC) = 0$

Now from Ex. 3-13, $(AB) + (AC) - (BC) \le (A)$

$\therefore \quad (A) + (AC) \le (A)$

$\therefore \quad (AC) \le 0$

$\therefore \quad (AC) = 0 \qquad (\because (AC) \ge 0)$

Ex. 3-21. *Given that* $(A) = (B) = (C) = \dfrac{N}{2}$ *and 80% of the A's are B's, 75% of A's are C's, find the limits to the percentage of B's that are C's.*

Sol. It is given that

$$(AB) = 0{\cdot}8\,(A) = 0{\cdot}4\,N$$
$$(AC) = 0{\cdot}75\,(A) = 0{\cdot}375\,N$$

and the limits of (BC) are to be obtained. From Ex. 3-13,

$(AB) + (AC) - (BC) \le (A)$

$\therefore \quad (BC) \ge (AB) + (AC) - (A) = 0{\cdot}4\,N + 0{\cdot}375\,N - 0{\cdot}5\,N$

$\therefore \quad (BC) \ge 0{\cdot}275\,N = 0{\cdot}55\,(B)$

and $\quad (AB) - (AC) + (BC) \le (B)$

$\therefore \quad (BC) \le (B) - (AB) + (AC) = 0{\cdot}5\,N - 0{\cdot}4\,N + 0{\cdot}375\,N$

$\therefore \quad (BC) \le 0{\cdot}475\,N = 0{\cdot}95(B)$

$$\therefore \quad 0{\cdot}55 \le \frac{(BC)}{(B)} \le 0{\cdot}95$$

$\therefore$ Reqd. limits are 55% and 95%.

Ex. 3-22. *Among the adult population of a certain town 50% of the population are male, 60% are wage-earners and 50% are 45 years of age or over, 10% of the males are not wage earners and 40% of the males are under 45. Can you infer anything about what percentage of the population of 45 or over are wage-earners?*

Sol. Let A, B, C denote the attributes of being male, wage-earner and 45 years old or more. Then

$N = 100,\ (A) = 50,\ (B) = 6$

The limits of (BC) are to be

Now $(AB) =$

and $(AC) =$

From Ex. 3-13, $(AB) + (AC$

$\therefore \quad (BC) \ge (AB) + (AC) -$

$\therefore \quad (BC) \ge 25$

and $\quad (AB) - (AC) + (BC) \le$

$\therefore \quad (BC) \le (B) + (AC) - ($

$\therefore \quad (BC) \le 45$

$\therefore \quad 25 \le (BC) \le 45.$

$\therefore$ Percentage of the pop

between $\dfrac{25}{50} \times 100 = 50\%$ and

Ex. 3-23. (*a*) *The following defects amongst a number of sc*

A = development defects, B

$N = 10{,}000,\ (A) = 877,\ (B)$

$(AB) = 338,\ (BC) = 455.$

Show that some dull boys a least must be so.

(*b*) *The following are the c*

$N = 10{,}000,\ (A) = 682,\ (B$

$(AB) = 248,\ (BC) = 363.$

Show that some defectively must be so.

Sol. (*a*) It is required to fi

Now $(\alpha C) = (C) - (AC)$

Now $(AC) + (AB) - (BC) \le$

and $\quad -(AB) + (AC) + (BC)$

$\therefore \quad (AC) \le (A) + (BC) - ($

$\therefore \quad (AC) \le 994$

and $\quad (AC) \le (C) + (AB) - ($

$\therefore \quad (AC) \le 672$

$\therefore \quad (\alpha C) \ge (C) - 672 = 7$

(*b*) Left as an exercise.

Ex. 3-24. *Given that 50% c 60), 80% non-able-bodied, 35% bodied and aged, find the great men.*

Sol. Let A, B and C denote

Then $N = 100,\ (A) = 50,\ ($

$(AB) = 35,\ (AC) = 4$

and the limits of (ABC) are to b

From Ex. 3-13, $\quad (ABC)$

(ABC)

smallpox 70% of the inhabitants
west percentage of the vaccinated

ced and vaccinated. Then

been attacked

rox.)

that all A's are C's.

hat no A's are C's.

f the A's are B's, 75% of A's are

5 N

5 N

wn 50% of the population are
over, 10% of the males are not
ou infer anything about what
s ?
age-earner and 45 years old or

$N = 100,\ (A) = 50,\ (B) = 60,\ (C) = 50,\ (A\beta) = \dfrac{10}{100} \times 50 = 5$

and $(A\gamma) = \dfrac{40}{100} \times 50 = 20$-

The limits of (BC) are to be obtained

Now $(AB) = (A) - (A\beta) = 50 - 5 = 45$

and $(AC) = (A) - (A\gamma) = 50 - 20 = 30$

From Ex. 3-13, $(AB) + (AC) - (BC) \le (A)$

$\therefore$ $(BC) \ge (AB) + (AC) - (A) = 45 + 30 - 50$

$\therefore$ $(BC) \ge 25$

and $(AB) - (AC) + (BC) \le (B)$

$\therefore$ $(BC) \le (B) + (AC) - (AB) = 60 + 30 - 45$

$\therefore$ $(BC) \le 45$

$\therefore$ $25 \le (BC) \le 45.$

$\therefore$ Percentage of the population of 45 years old or more who are wage-earners lies between $\dfrac{25}{50} \times 100 = 50\%$ and $\dfrac{45}{50} \times 100 = 90\%$.

Ex. 3-23. (*a*) *The following are the proportions of boys observed for certain classes of defects amongst a number of school-children.*

A = development defects, B = nerve signs, C = mental dullness.

$N = 10{,}000,\ (A) = 877,\ (B) = 1{,}086,\ (C) = 789$

$(AB) = 338,\ (BC) = 455.$

Show that some dull boys do not exhibit development defects and state how many at least must be so.

(*b*) *The following are the corresponding figures for girls :*

$N = 10{,}000,\ (A) = 682,\ (B) = 850,\ (C) = 689$

$(AB) = 248,\ (BC) = 363.$

Show that some defectively developed girls are not dull and state how many at least must be so.

Sol. (*a*) It is required to find the lower limit of (αC)

Now $(\alpha C) = (C) - (AC)$

Now $(AC) + (AB) - (BC) \le (A)$

and $-(AB) + (AC) + (BC) \le (C)$

$\therefore$ $(AC) \le (A) + (BC) - (AB) = 877 + 455 - 338$

$\therefore$ $(AC) \le 994$

and $(AC) \le (C) + (AB) - (BC) = 789 + 338 - 455$

$\therefore$ $(AC) \le 672$

$\therefore$ $(\alpha C) \ge (C) - 672 = 789 - 672 = 117$

(*b*) Left as an exercise.

Ex. 3-24. *Given that 50% of the inmates of an institution are men, 60% are aged (over 60), 80% non-able-bodied,* 35% aged men, 45% *non-able-bodied men and 42% non-able-bodied and aged, find the greatest and least possible proportions of non-able-bodied aged men.*

Sol. Let A, B and C denote the attributes of being man, aged and non-able-bodied.

Then $N = 100,\ (A) = 50,\ (B) = 60,\ (C) = 80.$

$(AB) = 35,\ (AC) = 45$ and $(BC) = 42$

and the limits of (ABC) are to be obtained.

From Ex. 3-13, $(ABC) \ge (AB) + (AC) - (A) = 30$

$(ABC) \ge (AB) + (BC) - (B) = 17$

and $\qquad (ABC) \geq (AC) + (BC) - (C) = 7$

$\therefore \quad (ABC) \geq 30$ satisfies all the three inequalities.

Also $(ABC) \leq (AB) = 35,\ (ABC) \leq (AC) = 45,\ (ABC) \leq (BC) = 42$

and $\quad (ABC) \leq (AB) + (AC) + (BC) - (A) - (B) - (C) + N = 32$

$\therefore \quad (ABC) \leq 32$

$\therefore \quad 30 \leq (ABC) \geq 32.$

Ex. 3-25. *50% of the imports of barley into a country come from the Dominions : 80% of the total imports go to brewing; 75% of the imports are grown in Northern Hemisphere; 80% of Northern-grown barley goes to brewing; 100% of foreign Southern-grown barley goes to stock-feeding. Show that the foreign Northern grown barley which goes to brewing cannot be less than 30% nor more than 50% of the total imports. (It is assumed that brewing and stock-feeding are the only two uses to which imported barley is put).*

Sol. Let A, B and C denote the attributes of the barley coming from dominions, being used in brewing and growing in Northern Hemisphere respectively. Then

$$N = 100,\ (A) = 50,\ (B) = 80,\ (C) = 75$$

$$(BC) = \frac{80}{100} \times 75 = 60 \text{ and } (\alpha\gamma) = (\alpha\beta\gamma)$$

and it is to be shown that

$$30 \leq (\alpha BC) \leq 50$$

Now $\qquad (\alpha\beta\gamma) = (\alpha\gamma) = (\alpha B\gamma) + (\alpha\beta\gamma)$

$\therefore \qquad (\alpha B\gamma) = 0$

or $\qquad (\alpha B) - (\alpha BC) = 0$ *i.e.*, $(\alpha B) = (\alpha BC)$

From Ex. 3-12, $(\alpha B) \geq (\alpha) + (B) - N = 30$

and $\qquad (\alpha B) \leq (\alpha) = 50 \qquad \because (\alpha) = N - (A) = 50$

$\therefore \qquad 30 \leq (\alpha BC) \leq 50.$

Ex. 3-26. *A penny is tossed three times and the results, heads and tails, noted. The process is continued until there are 100 sets of threes. In 69 cases heads fell first, in 49 cases heads fell second and in 53 cases heads fell third. In 33 cases heads fell both first and second and in 21 cases heads fell both second and third. Show that there must have been at least 5 occasions on which heads fell three times and that there could not have been more than 15 occasions on which tail fell three times, though there need not have been any.*

Sol. Let A, B and C denote the attributes of getting head in first, second and third trial respectively. Then

$N = 100,\ (A) = 69,\ (B) = 49,\ (C) = 53,\ (AB) = 33,\ (BC) = 21$

From Ex. 3-13, $(ABC) \geq (AB) + (BC) - (B) = 5$

$(\alpha\beta\gamma) \leq (\alpha\beta) = \alpha\beta.N = \{1 - A\} \cdot \{1 - B\}.N = N - (A) - (B) + (AB) = 15$

and $\quad (\alpha\beta\gamma) \leq (\beta\gamma) = N - (B) - (C) + (BC) = 19$

$(\alpha\beta\gamma) \leq 15.$

Ex. 3-27. *Given that* $(A) = (B) = (C) = \dfrac{N}{2}$ *and that*

$$\frac{(AB)}{N} = \frac{(AC)}{N} = p,$$

find what must be the greatest and least values of p in order that we may infer that $\dfrac{(BC)}{N}$ *exceeds any given value, say q.*

Sol. From Ex. 3-13, $(AB) + (BC) + (AC) \geq (A) + (B) + (C) - N$

$$\therefore \qquad 2Np + (BC) \geq \frac{N}{2}$$

$\therefore \qquad \dfrac{(BC)}{N} \geq \dfrac{1}{2}$ ·

and $\qquad (AB) + (AC) - (BC) \leq (A)$

$\therefore \qquad 2pN - (BC) \leq \dfrac{N}{2}$

$\therefore \qquad \dfrac{(BC)}{N} \geq 2p$

Since $\dfrac{(BC)}{N}$ is to exceed q, from

$\dfrac{1}{2} - 2p \geq q$ and 2

i.e., $\qquad p \leq \dfrac{1}{4}(1 - 2q)$ and p

From Ex. 3-12, $(AB) \leq (A)$

$\therefore \qquad p \leq \dfrac{1}{2}$

Since $\qquad (AB) \nless 0,$

$\therefore \quad p$ must lie between 0 and $\frac{1}{4}$

Ex. 3-28. *Show that if* $\dfrac{(A)}{N} = x$

and $\qquad \dfrac{(AB)}{N} = \dfrac{(A}{}$

the value of neither x nor y can excee

Sol. From Ex. 3-12, $(BC) \geq (B$

and $\qquad (AB) \leq (A$

$\therefore \qquad x \geq 5x$

$\therefore \qquad y \leq x \leq$

Ex. 3-29. *Show that for n attrib*

$(ABC \ldots\ldots M) \geq \{(A) + (B) + (C)$

Sol. From Ex. 3-12, $(AB) \geq (A$

Replacing B by BC

$(ABC) \geq (A$

Also $\qquad (BC) \geq (B$

$\therefore \qquad (ABC) \geq (A$

In general for m attributes $A, B,$

$(ABC \ldots P) \geq (A) + (B) +$

Replacing P by RS and using

$(RS) \geq (R$

for $(m + 1)$ attributes

$(ABC \ldots\ldots RS) \geq (A) + (B)$ ·

$\therefore \quad$ If the inequality holds for

the inequality holds for three attribute

of attributes.

$(BC) = 42$
$= 32$

me from the Dominions : 80%
own in Northern Hemisphere;
reign Southern-grown barley
barley which goes to brewing
ts. (It is assumed that brewing
rley is put).
oming from dominions, being
tively. Then

$\because (\alpha) = N - (A) = 50$

, heads and tails, noted. The
es heads fell first, in 49 cases
ses heads fell both first and
that there must have been at
re could not have been more
need not have been any.
in first, second and third trial

$= 21$

$A) - (B) + (AB) = 15$

that we may infer that $\dfrac{(BC)}{N}$

$C) - N$

$$\therefore \qquad \frac{(BC)}{N} \geq \frac{1}{2} - 2p \qquad \text{...(1)}$$

and $\qquad (AB) + (AC) - (BC) \leq (A)$

$$\therefore \qquad 2pN - (BC) \leq \frac{N}{2}$$

$$\therefore \qquad \frac{(BC)}{N} \geq 2p - \frac{1}{2} \qquad \text{...(2)}$$

Since $\dfrac{(BC)}{N}$ is to exceed q, from (1) and (2)

$$\frac{1}{2} - 2p \geq q \text{ and } 2p - \frac{1}{2} \geq q$$

i.e., $\qquad p \leq \frac{1}{4}(1 - 2q)$ and $p \geq \frac{1}{4}(2q + 1) \qquad$ ...(3)

From Ex. 3-12, $(AB) \leq (A)$

$$\therefore \qquad p \leq \frac{1}{2}$$

Since $\qquad (AB) \nless 0,\ p \geq 0$

$\therefore \quad p$ must lie between 0 and $\frac{1}{4}(1 - 2q)$ or between $\frac{1}{4}(1 + 2q)$ and $\frac{1}{2}$.

Ex. 3-28. *Show that if* $\dfrac{(A)}{N} = x = \dfrac{(B)}{2N} = \dfrac{(C)}{3N}$

and $$\frac{(AB)}{N} = \frac{(AC)}{N} = \frac{(BC)}{N} = y$$

the value of neither x nor y can exceed $\frac{1}{4}$.

Sol. From Ex. 3-12, $(BC) \geq (B) + (C) - N$ *i.e.,* $y \geq 5x - 1$

and $\qquad (AB) \leq (A)$ *i.e.,* $y \leq x$,

$$\therefore \qquad x \geq 5x - 1 \text{ i.e., } x \leq \frac{1}{4}$$

$$\therefore \qquad y \leq x \leq \frac{1}{4}.$$

Ex. 3-29. *Show that for n attributes A, B, C, ...M*
$(ABC \ldots\ldots M) \geq \{(A) + (B) + (C) + \ldots + (M)\} - (n-1)N$ *where N is the total frequency.*

Sol. From Ex. 3-12, $(AB) \geq (A) + (B) - N$

Replacing B by BC

$$(ABC) \geq (A) + (BC) - N$$

Also $\qquad (BC) \geq (B) + (C) - N$

$$\therefore \qquad (ABC) \geq (A) + (B) + (C) - 2N$$

In general for m attributes $A, B, C, \ldots P$, let

$$(ABC \ldots P) \geq (A) + (B) + (C) + \ldots + (P) - (m-1)N$$

Replacing P by RS and using

$$(RS) \geq (R) + (S) - N,$$

for $(m + 1)$ attributes

$$(ABC \ldots RS) \geq (A) + (B) + \ldots + (S) - mN$$

$\therefore$ If the inequality holds for m attributes it also holds for $(m + 1)$ attributes. Since the inequality holds for three attributes, it also holds for four and hence five and any number of attributes.

Ex. 3-30. *In a very hotly fought battle 70% at least of the combatants lost an eye, 75% at least lost an ear, 80% at least lost an arm and 85% at least lost a leg. How many at least must have lost all four ?*

Sol. Let A, B, C, D denote losing an eye, an ear, an arm and a leg respectively. Then $N = 100$, $(A) \geq 70$, $(B) \geq 75$, $(C) \geq 80$ and $(D) \geq 85$. Now from Ex. 3-29.

$$\begin{aligned}(ABCD) &\geq (A)+(B)+(C)+(D)-3N \\ &\geq 70+75+80+85-300 \\ &= 10\end{aligned}$$

$\therefore$ 10% at least have lost all four.

3.2. Association of Attributes

Independence. *If there is no relationship of any kind between two attributes A and B, it is expected to have the same proportion of A's among B's as among not B's i.e., β's. Two such attributes are termed as independent and the criterion of independence of two attributes A and B is*

$$\frac{(AB)}{(B)} = \frac{(A\beta)}{(\beta)}$$

Association. *If A and B are not independent, these are related in some way or other, however, complicated. These are said to be positively associated or simply associated if*

$$(AB) > \frac{(A)(B)}{N} \qquad \textit{negatively associated if} \qquad (AB) < \frac{(A)(B)}{N}$$

In statistics A and B are said to be associated only if these appear together in a greater number of cases than is to be expected if these are independent.

Complete Association. *Two attributes are said to be completely associated if one of them cannot occur without the other though the other may occur without the one i.e., all A's are B's or all B's are A's according as whether A's or B's are in minority.*

Complete Disassociation. *It may be taken either as the case when no A's are B's; or the case when no α's are β's.*

Co-efficient of Association. *It is given by*

$$Q = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$$

Co-efficient of Colligation. *It is defined by*

$$R = \frac{\sqrt{\{(AB)(\alpha\beta)\}}-\sqrt{\{(A\beta)(\alpha B)\}}}{\sqrt{\{(AB)(\alpha\beta)\}}+\sqrt{\{(A\beta)(\alpha B)\}}}$$

Symbols $\qquad (AB)_0 = \dfrac{(A)(B)}{N}$

$$\delta = (AB)-(AB)_0.$$

Ex. 3-31. *If A and B be two independent attributes, prove that*

$$(i)\ \frac{(\alpha B)}{(B)} = \frac{(\alpha\beta)}{(\beta)} \qquad (ii)\ \frac{(\alpha\beta)}{(\alpha)} = \frac{(A\beta)}{(A)} \qquad (iii)\ \frac{(AB)}{(A)} = \frac{(\alpha\beta)}{(\alpha)}$$

and (iv) $(AB) = \dfrac{(A)(B)}{N}$.

Sol. Since A and B are independent,

$$\frac{(AB)}{(B)} = \frac{(A\beta)}{(\beta)} \qquad \text{...(1)}$$

of the combatants lost an eye, 75%
least lost a leg. How many at least

arm and a leg respectively. Then N
rom Ex. 3-29.
$-3N$
00

d between two attributes A and B,
B's as among not B's i.e., β's. Two
n of independence of two attributes

are related in some way or other,
ociated or simply associated if

$$(AB) < \frac{(A)(B)}{N}$$

hese appear together in a greater
endent.
e completely associated if one of
occur without the one i.e., all A's
are in minority.
the case when no A's are B's; or

$)(\alpha B)\}$
$)(\alpha B)\}$

rove that

ii) $\dfrac{(AB)}{(A)} = \dfrac{(\alpha\beta)}{(\alpha)}$

...(1)

(*i*) From (1), $\dfrac{(AB)}{(A\beta)} = \dfrac{(B)}{(\beta)} = \dfrac{(AB)+(\alpha B)}{(A\beta)+(\alpha\beta)}$

$\therefore$ each ratio $= \dfrac{(\alpha B)}{(\alpha B)}$

$\therefore \dfrac{(B)}{(\beta)} = \dfrac{(\alpha B)}{(\alpha\beta)}$ or $\dfrac{(\alpha B)}{(B)} = \dfrac{(\alpha\beta)}{(\beta)}$

(*ii*) From (1), $\dfrac{(AB)}{(B)} = \dfrac{(A\beta)}{(\beta)} = \dfrac{(AB)+(A\beta)}{(B)+(\beta)} = \dfrac{(A)}{N}$ ...(2)

$\therefore \dfrac{(A\beta)}{(\beta)} = \dfrac{(A)}{N}$

$\therefore \dfrac{(A\beta)}{(\beta)-(A\beta)} = \dfrac{(A)}{N-(A)}$ or $\dfrac{(A\beta)}{(\alpha\beta)} = \dfrac{(A)}{(\alpha)}$

$\therefore \dfrac{(A\beta)}{(A)} = \dfrac{(\alpha\beta)}{(\alpha)}$

(*iii*) From (2), $\dfrac{(AB)}{(B)} = \dfrac{(A)}{N}$

$\therefore \dfrac{(AB)}{(B)-(AB)} = \dfrac{(A)}{N-(A)}$ or $\dfrac{(AB)}{(\alpha B)} = \dfrac{(A)}{(\alpha)}$

$\therefore \dfrac{(AB)}{(A)} = \dfrac{(\alpha B)}{(\alpha)}$

(*iv*) From (2), $(AB) = \dfrac{(A)(B)}{N}$.

Ex. 3-32. *Show, whether A and B are independent, positively associated or negatively associated in each of the following cases :*

(*i*) $N = 100$, $(A) = 47$, $(B) = 62$, $(AB) = 32$.
(*ii*) $(A) = 490$, $(AB) = 294$, $(\alpha) = 570$, $(\alpha B) = 380$.
(*iii*) $(AB) = 256$, $(\alpha B) = 768$, $(A\beta) = 48$, $(\alpha\beta) = 144$.

Sol. (*i*) $\dfrac{(A)(B)}{N} = \dfrac{(47)(62)}{100} = 29{\cdot}14 = (AB)_0$

Since $(AB) > (AB)_0$, attributes are positively associated.

(*ii*) $N = (\alpha) + (A) = 1060$, $(B) = (AB) + (\alpha B) = 674$.

$\therefore (AB)_0 = \dfrac{(A)(B)}{N} = \dfrac{(490)(674)}{1060} = 311{\cdot}6$ (nearly)

$\therefore (AB) < (AB)_0$

$\therefore$ Attributes are negatively associated.

(*iii*) $(A) = (AB) + (A\beta) = 256 + 48 = 304$
$(B) = (AB) + (\alpha B) = 256 + 768 = 1024$
$(\alpha) = (\alpha B) + (\alpha\beta) = 768 + 144 = 912$

$\therefore N = (A) + (\alpha) = 304 + 912 = 1216$

$\therefore (AB)_0 = \dfrac{(A)(B)}{N} = \dfrac{(304)(1024)}{1216} = 256 = (AB)$

$\therefore$ A and B are independent.

Ex. 3-33. *The male population of U.P. is 250 lakhs. The number of literate males is 20 lakhs and the total number of male criminals is 26 thousands. The number of literate male*

criminals is two thousands. Do you find any association between literary and criminality ?

Sol. Let A and B denote the attributes of being literate and criminal respectively. Then

$(A) = 20$ lakhs, $(B) = 26$ thousands $= 0{\cdot}26$ lakhs

$(AB) = 2$ thousands $= 0{\cdot}02$ lakhs and $N = 250$ lakhs

$$\therefore \quad (AB)_0 = \frac{(A)(B)}{N} = \frac{(20)(0{\cdot}26)}{250} = 0{\cdot}0208 > (AB)$$

$\therefore$ A and B are negatively associated.

Ex. 3.34. *Show that*

(*i*) $Q = \dfrac{N\delta}{(AB)\,(\alpha\beta) + (A\beta)(\alpha B)}$ *and deduce that* $Q = 0$ *when* A *and* B *are independent.*

(*ii*) $Q = 1$ *for complete association.*

(*iii*) $Q = -1$ *for complete disassociation.*

(*iv*) $Q = \dfrac{2R}{1+R^2}$ *where* R *is the co-efficient of colligation.*

Sol. (*i*)

$$\delta = (AB) - (AB)_0 = (AB) - \frac{(A)(B)}{N}$$

$$\begin{aligned} \therefore \quad N\delta &= (AB)\,\{(A) + (\alpha)\} - (A)\,\{(AB) + (\alpha B)\} \\ &= (AB)\,\{(\alpha B) + (\alpha\beta) - \{(AB) + (A\beta)\}\,(\alpha B) \\ &= (AB)\,(\alpha\beta) - (A\beta)\,(\alpha B) \end{aligned}$$

$$\therefore \quad Q = \frac{N\delta}{(AB)\,(\alpha\beta) + (A\beta)(\alpha B)}$$

If A and B are independent, $\delta = 0$

$$\therefore \quad Q = 0$$

(*ii*) If there is complete association, all A's are B's or all B's are A's according as A's or B's are in minority.

$\therefore$ Either $(A\beta) = 0$ or $(\alpha B) = 0$

$$\therefore \quad Q = 1.$$

(*iii*) If there is complete disassociation, either no A's are B's or no α's and β's.

$\therefore$ Either $(AB) = 0$ or $(\alpha\beta) = 0$

$$\therefore \quad Q = -1.$$

(*iv*) By def., $R = \dfrac{\sqrt{\{(AB)(\alpha\beta)\}} - \sqrt{\{(A\beta)(\alpha B)\}}}{\sqrt{\{(AB)(\alpha\beta)\}} + \sqrt{\{(A\beta)(\alpha B)\}}}$

$$\therefore \quad \frac{1+R}{1-R} = \frac{\sqrt{(AB)(\alpha\beta)}}{\sqrt{(A\beta)(\alpha B)}}$$

$$\therefore \quad Q = \left\{\frac{\dfrac{(1+R)^2}{(1-R)^2} - 1}{\dfrac{(1+R)^2}{(1-R)^2} + 1}\right\} = \frac{2R}{1+R^2}$$

Note. From (*i*) $\quad Q > 0$ for positive association
$\quad < 0$ for negative association
$\quad = 0$ for independence.

Ex. 3-35. *If A and B are independent, find the* (AB), $(A\beta)$, (αB) *and* $(\alpha\beta)$.

Sol. Since A and B are independent, $(AB) = \dfrac{(A)(B)}{N}$

$\therefore \quad (A\beta) =$

$=$

$(\alpha B) =$

$(\alpha\beta) =$

$=$

$=$

$=$

Ex. 3-36. *Show that* $\delta = \dfrac{(B}{}$

Sol. $\delta =$

$=$

$=$

$=$

Interchanging A and B

$\delta =$

Ex. 3-37. *Show that*

$(AB)^2 + (\alpha\beta)^2 - (\alpha B)^2 -$

Sol. R.H.S. $= \{(A) - (\alpha)\}$

$= [\{(AB) + (A$

$- \{(A\beta) + ($

$= \{(AB) - (\alpha$

$= \{(AB) - (\alpha$

$= \{(AB) - (\alpha$

$+ 2\,\{(\alpha B) -$

$= (AB)^2 + (\alpha$

Ex. 3-38. *Show that if*

$(A$

$(A$

be two aggregates corresponding

$(AB)_1 - (AB)_2 = (\alpha B)_2 - (\alpha B)$

Sol. $(A) =$

$$\therefore \quad (A\beta) = (A) - (AB) = (A)\left\{1 - \frac{(B)}{N}\right\} = \frac{(A)\{N-(B)\}}{N}$$
$$= \frac{(A)(\beta)}{N}$$
$$(\alpha B) = (B) - (AB) = (B)\left\{1 - \frac{(A)}{N}\right\} = \frac{(\alpha)(B)}{N}$$
$$(\alpha\beta) = N - (A) - (B) + (AB)$$
$$= N - (A) - (B) + \frac{(A)(B)}{N}$$
$$= (\alpha) - \frac{(B)}{N}\{N-(A)\}$$
$$= (\alpha)\left\{1 - \frac{(B)}{N}\right\} = \frac{(\alpha)(\beta)}{N}.$$

Ex. 3-36. *Show that* $\delta = \dfrac{(B)(\beta)}{N}\left\{\dfrac{(AB)}{(B)} - \dfrac{(A\beta)}{(\beta)}\right\}.$

Sol.
$$\delta = (AB) - \frac{(A)(B)}{N} = \frac{1}{N}\{N(AB) - (A)(B)\}$$
$$= \frac{1}{N}[(AB)\{(B)+(\beta)\} - \{(AB)+(A\beta)\}(B)]$$
$$= \frac{1}{N}[(AB)(\beta) - (A\beta)(B)]$$
$$= \frac{(B)(\beta)}{N}\left[\frac{(AB)}{(B)} - \frac{(A\beta)}{(\beta)}\right]$$

Interchanging A and B
$$\delta = \frac{(A)(\alpha)}{N}\left[\frac{(AB)}{(A)} - \frac{(\alpha B)}{(\alpha)}\right]$$

Ex. 3-37. *Show that*
$$(AB)^2 + (\alpha\beta)^2 - (\alpha B)^2 - (A\beta)^2 = [(A)-(\alpha)]\,[(B)-(\beta)] + 2N\delta.$$

Sol. R.H.S.
$$= \{(A)-(\alpha)\}\{(B)-(\beta)\} + 2N\left\{(AB) - \frac{(A)(B)}{N}\right\}$$
$$= [\{(AB)+(A\beta)\} - \{(\alpha B)+(\alpha\beta)\}]\,[\{(AB)+(\alpha B)\} - \{(A\beta)+(\alpha\beta)\}] + 2N(AB) - 2(A)(B)$$
$$= \{(AB)-(\alpha\beta)\}^2 - \{(A\beta)-(\alpha B)\}^2 + 2\{(A)+(\alpha)\}(AB) - 2(A)(B)$$
$$= \{(AB)-(\alpha\beta)\}^2 - \{(A\beta)-(\alpha B)\}^2 + 2(A)\{(AB)-(B)\} + 2(\alpha)(AB)$$
$$= \{(AB)-(\alpha\beta)\}^2 - \{(A\beta)-(\alpha B)\}^2 - 2\{(AB)+(A\beta)\}(\alpha B) + 2\{(\alpha B)+(\alpha\beta)\}(AB)$$
$$= (AB)^2 + (\alpha\beta)^2 - (A\beta)^2 - (\alpha B)^2.$$

Ex. 3-38. *Show that if*
$$(AB)_1,\ (\alpha B)_1,\ (A\beta)_1,\ (\alpha\beta)_1$$
$$(AB)_2,\ (\alpha B)_2,\ (A\beta)_2,\ (\alpha\beta)_2$$
be two aggregates corresponding to the same values of (A), (B), (α) and (β)
$$(AB)_1 - (AB)_2 = (\alpha B)_2 - (\alpha B)_1 = (A\beta)_2 - (A\beta)_1 = (a\beta)_1 - (a\beta)_2.$$

Sol.
$$(A) = (AB)_1 + (A\beta)_1 = (AB)_2 + (A\beta)_2$$

$$\therefore \quad (AB)_1 - (AB)_2 = (A\beta)_2 - (A\beta)_1 \qquad ...(1)$$

$$(B) = (AB)_1 + (\alpha B)_1 = (AB)_2 + (\alpha B)_2$$

$$\therefore \quad (AB)_1 - (AB)_2 = (\alpha B)_2 - (\alpha B)_1 \qquad ...(2)$$

$$(\alpha) = (\alpha B)_1 + (\alpha\beta)_1 = (\alpha B)_2 + (\alpha\beta)_2$$

$$\therefore \quad (\alpha B)_1 - (\alpha B)_2 = (\alpha\beta)_2 - (\alpha\beta)_1 \qquad ...(3)$$

From (1), (2) and (3) result follows.

Ex. 3-39. *Investigate the association between eye colour of husband and eye-colour of wife from the data given below :*

Husbands with light eyes and wives with light eyes = 309
Husbands with light eyes and wives with not light eyes = 214
Husbands with not light eyes and wives with light eyes = 132
Husbands with not light eyes and wives with not light eyes = 119

Sol. Let A, B denote the attributes of husbands with light eyes and wives with light eyes repectively.

Then $(AB) = 309$, $(A\beta) = 214$, $(\alpha B) = 132$ and $(\alpha\beta) = 119$.

$$\therefore \quad Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} = \frac{(309)(119) - (214)(132)}{(309)(119) + (214)(132)}$$

$$= \frac{36771 - 28248}{36771 + 28248} = 0{\cdot}13$$

$\therefore$ There seems to be positive association of small degree. Working out percentage :

Percentages of light-eyed amongst the wives of light-eyed husbands $= \dfrac{309}{214+309} \times 100 = 59\%$ and percentage of light-eyed amongst the wives of not light-eyed husbands $= \dfrac{132}{132+119} \times 100 = 53\%$. Comparison brings out that the association is small, so small that no stress can be laid on it as indicating anything but a fluctuation of sampling.

Ex. 3-40. *Investigate the association between darkness of eye colour in father and son from the following data :*

Father with dark eyes and sons with dark eyes = 50
Father with dark eyes and sons without dark eyes = 79
Father without dark eyes and sons with dark eyes = 89
Father without dark eyes and sons without dark eyes = 782

What would have been the frequency of 'fathers with dark eyes and sons with dark eyes' for the same total number, had there been complete independence ?

Sol. Let A and B be the attributes of father and son to be with dark eyes respectively. Then

$(AB) = 50$, $(A\beta) = 79$, $(\alpha B) = 89$ and $(\alpha\beta) = 782$.

$$\therefore \quad Q = \frac{(50)(782) - (79)(89)}{(50)(782) + (79)(89)} = \frac{32069}{46131} = 0{\cdot}7$$

$\therefore$ There is a positive association of high degree between the darkness of eye-colour in father and son.

Now
$$(A) = (AB) + (A\beta) = 50 + 79 = 129$$
$$(B) = (AB) + (\alpha B) = 50 + 89 = 139$$
$$N = 50 + 79 + 89 + 782 = 1000$$

$$\therefore \quad (AB)_0 = \frac{(A)(B)}{N} = \frac{(129)(139)}{1000} = 18 \text{ (approx.)}$$

$\therefore$ Had A and B been independent.

$(AB) = (Al$

Ex. 3-41. *The following data rel persons. You are required to calcula and unemployment and interpret it.*

Illiterate unemployed
Literate employed
Illiterate employed

Sol. Let A and B denote the attr

Then

$(AB) = 50$
$(A\beta) = 20$

$\therefore \quad Q =$

$= 0\cdot$

Thus there is a positive associati

Thus in general literate person in un

Ex. 3-42. *From the following, f*

Total population = 16,264,000, 7,623, *number of bald-headed blind*

Sol. Let A and B denote the a

Then

$(AB) = 2$

$\therefore \quad (AB)_0 =$

$\therefore \quad (AB) > ($

$\therefore$ There is positive associatio

Ex. 3-43. *Do you find any asso the following data :*

Good-natured brothers and go
Good-natured brothers and sul
Sullen-natured brothers and go
Sullen-natured brothers and su

Sol. Let A and B denote the respectively. Then

$(AB) = 1230$, $(A\beta) = 850$, $(\alpha B$

$\therefore \quad Q =$

$= ($

$\therefore$ There is positive associa

Ex. 3-44. *Can vaccination b the data given below ?*

'Of 1482 persons in a locality

'Of 1482 persons, 343 had be

Sol. Let A and B denote the

Then

$N = 1482$, $(AB) = 35$, $($

$$(AB) = (AB)_0 = 18.$$

Ex. 3-41. *The following data relates to literacy and unemployment in a group of 500 persons. You are required to calculate Yule's co-efficient of association between literacy and unemployment and interpret it.*

Illiterate unemployed	220
Literate employed	20
Illiterate employed	180

Sol. Let A and B denote the attributes of being literate and unemployed respectively. Then

$$(AB) = 500 - (220 + 20 + 180) = 80$$
$$(A\beta) = 20 \quad (\alpha B) = 220 \text{ and } (\alpha\beta) = 180$$
$$\therefore \quad Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} = \frac{(80)(180) - (20)(220)}{(80)(180)+(20)(220)}$$
$$= 0{\cdot}5.$$

Thus there is a positive association of high degree between literacy and unemployment. Thus in general literate person in unemployed.

Ex. 3-42. *From the following, find whether blindness and baldness are associated.*

Total population = 16,264,000, *number of bald-headed* = 24,441, *number of blind* = 7,623, *number of bald-headed blind* = 221.

Sol. Let A and B denote the attributes of being bald-headed and blind respectively. Then

$$(AB) = 221, \ (A) = 24{,}441, \ (B) = 7{,}623 \text{ and } N = 16{,}264{,}000$$
$$\therefore \quad (AB)_0 = \frac{(A)(B)}{N} = \frac{(24{,}441)(7{,}623)}{16{,}264{,}000} \simeq 11$$
$$\therefore \quad (AB) > (AB)_0$$

$\therefore$ There is positive association between baldness and blindness.

Ex. 3-43. *Do you find any association between the tempers of brothers and sisters from the following data :*

Good-natured brothers and good-natured sisters = 1230
Good-natured brothers and sullen-natured sisters = 850
Sullen-natured brothers and good-natured sisters = 530
Sullen-natured brothers and sullen-natured sisters = 980

Sol. Let A and B denote the attributes of being good-natured for brother and sister respectively. Then

$(AB) = 1230, \ (A\beta) = 850, \ (\alpha B) = 530$ and $(\alpha\beta) = 980$

$$\therefore \quad Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} = \frac{(1230)(980) - (850)(530)}{(1230)(980) + (850)(530)}$$
$$= 0{\cdot}46.$$

$\therefore$ There is positive association.

Ex. 3-44. *Can vaccination be regarded as a preventive measure for small-pox from the data given below ?*

'*Of* 1482 *persons in a locality exposed to small pox,* 368 *in all were attacked*'.

'*Of* 1482 *persons,* 343 *had been vaccinated and of these only* 35 *were attacked*'.

Sol. Let A and B denote the attributes of being vaccinated and attacked respectively. Then

$$N = 1482, \ (AB) = 35, \ (A) = 343, \ (B) = 368.$$

$$(AB)_0 = \frac{(A)(B)}{N} = \frac{(343)(368)}{1482} = 85{\cdot}2$$

$$\therefore \quad (AB) < (AB)_0$$

$\therefore$ There is negative association and hence vaccination can be regarded as a preventive measure for small-pox.

Ex. 3-45. *In an antimalarial campaign in a certain area, quinine was administered to* 812 *persons out of a total population of* 3248. *The number of fever cases is shown below :*

Treatment	*Fever*	*No Fever*
Quinine	20	792
No quinine	220	2,216

Discuss the usefulness of quinine in checking malaria.

Sol. Let A and B correspond to fever and quinine. Then

$$(AB) = 20,\ (\alpha B) = 792,\ (A\beta) = 220 \text{ and } (\alpha\beta) = 2{,}216$$

$$\therefore \quad Q = \frac{(20)(2216) - (792)(220)}{(20)(2216) + (792)(220)} = -\,0{\cdot}6$$

$\therefore$ There is a negative association of high degree. Hence quinine may be taken as preventive malaria.

Ex. 3-46. *A group of* 1000 *fathers was studied and it was found that* 12·9% *had dark eyes. Among them the ratio of those having sons with dark eyes to those having sons with not dark eyes was* 1 : 1·58. *The number of cases where fathers and sons both did not have dark eyes was* 782. *Calculate a co-efficient of association between darkness of eye-colour in father and son. Give the frequencies that would have been observed had there been completely no heredity.*

Sol. Let A and B denote the attributes of having dark eyes for father and son respectively. Then

$$(A) = \frac{12{\cdot}9}{100} \times 1000 = 129,\ N = 1000,\ \frac{(AB)}{(A\beta)} = \frac{1}{1{\cdot}58} \text{ and } (\alpha\beta) = 782$$

$$\text{Now} \quad \frac{(AB)}{1} = \frac{(A\beta)}{1{\cdot}58} = \frac{(AB)+(A\beta)}{2{\cdot}58} = \frac{(A)}{2{\cdot}58} = \frac{129}{2{\cdot}58} = 50$$

$$\begin{aligned}
\therefore \quad (AB) &= 50 \text{ and } (A\beta) = 79 \\
(\beta) &= (A\beta) + (\alpha\beta) = 79 + 782 = 861 \\
\therefore \quad (B) &= N - (\beta) = 1000 - 861 = 139 \\
\therefore \quad (AB) + (\alpha B) &= 139 \\
\therefore \quad (\alpha B) &= 89
\end{aligned}$$

Now see Ex. 3-40.

EXERCISES

1. If a collection contains N items, each of which is characterized by one or more of the attributes A, B, C and D, show that with the usual notation

   (*i*) $(ABCD) \geq (A) + (B) + (C) + (D) - 3N$

   and (*ii*) $(ABCD) = (ABD) + (ACD) - (AD) + (AD\beta\gamma)$

   where β and γ represent the characteristics of the absence of B and C respectively.

2. Three aptitude tests A, B, C were given to 200 apprentice trainees. From amongst them 80 passed test A, 78 passed test B and 96 passed the third test. While 20 passed all three tests, 42 failed all the three, 18 passed A and B but failed C and 38 failed A

and B but passed the third. I
three tests and (*ii*) whether

3. In a college 50% of the stud
80% receive scholarships,
are boys receiving scholarsl
Determine the limits to the
scholarships.
4. A study was made about th
the following summary is g
'Of the students surveye
60% were irregular in their
thirds were from well-to-d
do families were 8. Is ther
5. The following data relate t
the two attributes are inde

Selling ability
Good
Poor

6. Calculate the co-efficient
data given below and inte
The total population of a ci
of criminals in the group c
7. The following data were c
Flowers violet, fruits prick
Flowers violet, fruits smo
Flowers white, fruits prick
Flowers white, fruits smo
Investigate the associatio
8. From the data given b
unemployment in the rura

Total number of adult ma
Literal males
Unemployed
Literate and unemployed

9. In an assortative mating s
the following information
was published.

Tall wives
Short wives

) = 85·2

on can be regarded as a preventive

area, quinine was administered to
er of fever cases is shown below :

No Fever
792
2,216

a.
hen
) = 2,216

$\frac{20)}{20)} = -0{\cdot}6$

Hence quinine may be taken as

t was found that 12·9% had dark
k eyes to those having sons with
thers and sons both did not have
between darkness of eye-colour
been observed had there been

es for father and son respectively.

and $(\alpha\beta) = 782$

$\frac{(A)}{2\cdot58} = \frac{129}{2\cdot58} = 50$

= 861
139

acterized by one or more of the
tation

γ)
ence of B and C respectively.
entice trainees. From amongst
the third test. While 20 passed
B but failed C and 38 failed A

and B but passed the third. Determine (*i*) how many trainees passed at least two of the three tests and (*ii*) whether the performances in tests A and B are associated.
[**Ans.** 76, $Q = 0{\cdot}3$]

3. In a college 50% of the students are boys, 60% of the students are above 18 years, and 80% receive scholarships, 35% of the students are boys above 18 years of age, 45% are boys receiving scholarships and 42% are above 18 years and receive scholarships. Determine the limits to the proportion of boys above 18 years who are in receipt of scholarships. [**Ans.** Lies between 30 and 32]
4. A study was made about the studying habits of the students of a certain university and the following summary is given at one place in the report.
 'Of the students surveyed 75% were from well-to-do families, 55% were boys and 60% were irregular in their studies. Out of the irregular one 50% were boys and two-thirds were from well-to-do families. The percentage of irregular boys from well-to-do families were 8. Is there any inconsistency in the data ? [**Ans.** Yes]
5. The following data relate to flexibility and selling ability of 20 salesmen. Test whether the two attributes are independent.

	Flexibility	
Selling ability	*Good*	*Poor*
Good	7	3
Poor	2	8

[**Ans.** $Q = 0{\cdot}8$]

6. Calculate the co-efficient of association between illiteracy and criminality from the data given below and interpret it.
 The total population of a city is 244,000 out of which 40,000 are literates. The number of criminals in the group of literates is 300 and in the group of illiterates 4,000.
[**Ans.** 0·5]
7. The following data were observed for hybrids of Datura.
 Flowers violet, fruits prickly $(AB) = 47$
 Flowers violet, fruits smooth $(A\beta) = 12$
 Flowers white, fruits prickly $(\alpha B) = 21$
 Flowers white, fruits smooth $(\alpha\beta) = 3$
 Investigate the association between colour of flower and character of fruit.
[**Ans.** (– 0·28)]
8. From the data given below, compare the association between literacy and unemployment in the rural and urban areas.

	Rural	*Urban*
Total number of adult males	25 lakhs	200 lakhs
Literal males	10 lakhs	40 lakhs
Unemployed	5 lakhs	4 lakhs
Literate and unemployed males	3 lakhs	4 lakhs

9. In an assortative mating study to find whether tall husbands tend to marry tall wives the following information about the wives of 125 tall and 125 short-statured husbands was published.

	Tall husbands (percent)	*Short husbands (percent)*
Tall wives	56	13
Short wives	11	48

Find the co-efficient of association between the stature of wives and husbands, ignoring medium-sized wives. **[Ans.** 0·9]

10. From the figures in the following table compare the association between literacy and unemployment in rural and urban areas.

	Urban	*Rural*
Total adult males	25 lakhs	20 lakhs
Literal males	10 lakhs	10 lakhs
Unemployed males	5 lakhs	12 lakhs
Literate and unemployed males	4 lakhs	4 lakhs

11. In a state with a total population of 70,000 adults, 34,000 are males and out of a total of 6,000 graduates 700 are females. Out of 1200 graduate employees of the state, 200 are females. Is there any sex bias in education among the people ? The state holds that no distinction is made in appointments in respect of sex. How far is their claim substantiated by the data given above ?

12. A census revealed the following figures of the blind and the insane in two age groups in a certain population.

	Age-group (15-25 years)	*Age-group* (over 25 years)
Total population	270,000	160,200
No. of blind	1,000	2,000
No. of insane	6,000	1,000
No. of insane among the blind	19	9

(*a*) Obtain a measure of the association between blindness and insanity in each of the two age group, (*b*) Do you consider that blindness and insanity are associated or disassociated with each other in the two age groups or more in one age group than the other ?

13. Obtain the co-efficient of association between unemployment and educational attainments from the following results of an urban survey.

	Employed	*Unemployed*
Illiterate or below matric	5997	432
Matric and above	572	96

[Ans. 0·4]

14. The following table gives the number of literates and criminals in three cities of U.P.

	Kanpur	*Allahabad*	*Agra*
Total number (in thousands)	244	184	230
Literates (in thousands)	30	47	33
Literate criminals (in thousands)	3	2	2
Illiterate criminals (in thousands)	40	20	24

Compare the degree of association between criminality and illiteracy in each of the above three cities.

□□

Difference Ope

4.1. Divided Differences

Let the values of $f(x)$ for $x = x_i$

$$f(x_0, x_1) =$$

$$f(x_0, x_1, x_2) =$$

$$f(x_0, x_1, x_2, x_3) =$$

and so on, are called divided diffe

Ex. 4-1. *Compute the divide*

x : 1

$f(x)$: 2

Sol.

x	$f(x)$	*1st order*
1	2	$\frac{4-2}{2-1} = 2$
2	4	$\frac{8-4}{3-2} = 4$
3	8	$\frac{16-8}{4-3} = 8$
4	16	$\frac{128-16}{7-4} =$
7	128	

Ex. 4-2. *Obtain the dividec*

Sol. Let $x_0, x_1 \ldots\ldots x_n$ be the

$$f(x_0, x_1) =$$

$$f(x_0, x_1, x_2) =$$

ature of wives and husbands, [**Ans.** 0·9]

sociation between literacy and

	Rural
	20 lakhs
	10 lakhs
	12 lakhs
	4 lakhs

)0 are males and out of a total te employees of the state, 200 people ? The state holds that sex. How far is their claim

l the insane in two age groups

	Age-group (over 25 years)
	160,200
	2,000
	1,000
	9

ss and insanity in each of the d insanity are associated or ore in one age group than the

nployment and educational ey.

ed	*Unemployed*
	432
	96

[**Ans.** 0·4]

minals in three cities of U.P.

Allahabad	*Agra*
184	230
47	33
2	2
20	24

and illiteracy in each of the

□□

4

Difference Operators and Interpolation

4.1. Divided Differences

Let the values of $f(x)$ for $x = x_0, x_1, \ldots\ldots x_n$ be known. Then

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0},$$

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

$$f(x_0, x_1, x_2, x_3) = \frac{f(x_1, x_2, x_3) - f(x_0, x_1, x_2)}{x_3 - x_0}$$

and so on, are called divided differences of first order, second order, third order, etc.

Ex. 4-1. *Compute the divided differences of $f(x)$ from the following table :*

x	:	1	2	3	4	7
$f(x)$	:	2	4	8	16	128

Sol.

x	$f(x)$	*1st order*	*2nd order*	*3rd order*	*4th order*
1	2	$\frac{4-2}{2-1} = 2$	$\frac{4-2}{3-1} = 1$		
2	4	$\frac{8-4}{3-2} = 4$	$\frac{8-4}{4-2} = 2$	$\frac{2-1}{4-1} = \frac{1}{3}$	$\frac{\frac{16}{15} - \frac{1}{3}}{7-1} = \frac{11}{90}$
3	8	$\frac{16-8}{4-3} = 8$	$\frac{\frac{112}{3} - 8}{7-3} = \frac{22}{3}$	$\frac{\frac{22}{3} - 2}{7-2} = \frac{16}{15}$	
4	16	$\frac{128-16}{7-4} = \frac{112}{3}$			
7	128				

Ex. 4-2. *Obtain the divided differences of $f(x) = x^2$.*

Sol. Let $x_0, x_1 \ldots\ldots x_n$ be the value of x. Then

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{x_1^2 - x_0^2}{x_1 - x_0} = x_1 + x_0$$

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

$$= \frac{(x_2+x_1)-(x_1+x_0)}{x_2-x_0} = 1$$

$$f(x_0, x_1, x_2, x_3) = \frac{f(x_1,x_2,x_3)-f(x_0,x_1,x_2)}{x_3-x_0} = 0$$

Evidently all higher order divided differences will be zero.

Ex. 4-3. *Prove that the divided differences are symmetrical in their arguments.*

Sol. Let $x_0, x_1, \ldots\ldots x_n$ be the arguments.

Then $$f(x_0, x_1) = \frac{f(x_1)-f(x_0)}{x_1-x_0} = \frac{f(x_1)}{x_1-x_0} + \frac{f(x_0)}{x_0-x_1}$$

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2)-f(x_0,x_1)}{x_2-x_0}$$

$$= \frac{1}{x_2-x_0}\left[\left\{\frac{f(x_2)}{x_2-x_1}+\frac{f(x_1)}{x_1-x_2}\right\}-\left\{\frac{f(x_1)}{x_1-x_0}+\frac{f(x_0)}{x_0-x_1}\right\}\right]$$

$$= \frac{f(x_2)}{(x_2-x_0)(x_2-x_1)} + \frac{f(x_1)}{(x_1-x_0)(x_1-x_2)} + \frac{f(x_0)}{(x_0-x_1)(x_0-x_2)}$$

Let in general

$$f(x_0, x_1, \ldots x_m) = \frac{f(x_m)}{(x_m-x_0)(x_m-x_1)\ldots\ldots(x_m-x_{m-1})}$$
$$+ \frac{f(x_{m-1})}{(x_{m-1}-x_0)(x_{m-1}-x_1)\ldots\ldots(x_{m-1}-x_{m-2})(x_{m-1}-x_m)}$$
$$+ \ldots\ldots + \frac{f(x_0)}{(x_0-x_1)(x_0-x_2)\ldots\ldots(x_0-x_m)}$$

Then

$$f(x_0, x_1, \ldots x_{m+1}) = \frac{f(x_1,x_2\ldots\ldots x_{m+1})-f(x_0,x_1\ldots\ldots x_m)}{x_{m+1}-x_0}$$

$$= \frac{1}{x_{m+1}-x_0}\left[\frac{f(x_{m+1})}{(x_{m+1}-x_1)(x_{m+1}-x_2)\ldots\ldots(x_{m+1}-x_m)}\right.$$
$$+ \frac{f(x_m)}{(x_m-x_1)(x_m-x_2)\ldots\ldots(x_m-x_{m-1})(x_m-x_{m+1})}$$
$$+ \ldots\ldots + \frac{f(x_1)}{(x_1-x_2)\ldots\ldots(x_1-x_{m+1})}$$
$$- \frac{1}{x_{m+1}-x_0}\left[\frac{f(x_m)}{(x_m-x_0)\ldots\ldots(x_m-x_{m-1})}\right.$$
$$+ \frac{f(x_{m-1})}{(x_{m-1}-x_0)\ldots\ldots(x_{m-1}-x_{m-2})(x_{m-1}-x_m)}$$
$$\left.+ \ldots\ldots + \frac{f(x_0)}{(x_0-x_1)\ldots\ldots(x_0-x_m)}\right]$$

$$= \frac{f(x_{m+1})}{(x_{m+1}-x_0)\ldots.(x_{m+1}-x_m)}$$

$$+ \frac{f(x_m)}{(x_m-x_0)\ldots.(x_m-x_{m-1})(x_m-x_{m+1})} + \ldots.$$

$$+ \frac{f(x_0)}{(x_0-x_1)\ldots.(x_0-x_{m+1})}$$

$\therefore$ By induction,

$$f(x_0, x_1, \ldots. x_n) = \frac{f(x_n)}{(x_n-x_0)\ldots.(x_n-x_{n-1})}$$

$$+ \frac{f(x_{n-1})}{(x_{n-1}-x_0)\ldots..(x_{n-1}-x_{n-2})(x_{n-1}-x_n)} + \ldots.$$

$$+ \frac{f(x_0)}{(x_0-x_1)\ldots.(x_0-x_n)}.$$

Evidently $f(x_0, x_1, \ldots. x_n)$ remains unchanged on interchanging the arguments. Hence $f(x_0, x_1, \ldots x_n)$ is symmetrical in its arguments.

Ex. 4-4. *Show that divided differences of the sum of two functions are equal to the sum of the divided differences of two functions.*

Sol. Let $f(x)$ and $g(x)$ be two f^ns and

$$h(x) = f(x) + g(x)$$

Now $$h(x_0, x_1 \ldots x_n) = \frac{h(x_n)}{\prod\limits_{i \neq n}(x_n - x_i)} + \frac{h(x_{n-1})}{\prod\limits_{i \neq n-1}(x_{n-1} - x_i)} + \cdots + \frac{h(x_0)}{\prod\limits_{i \neq 0}(x_0 - x_i)}$$

$$= \frac{1}{\prod\limits_{i \neq n}(x_n - x_i)}\{f(x_n) + g(x_n)\} + \frac{1}{\prod\limits_{i \neq n-1}(x_{n-1} - x_i)}\{f(x_{n-1}) + g(x_{n-1})\}$$

$$+ \ldots. + \frac{1}{\prod\limits_{i \neq 0}(x_0 - x_i)}\{f(x_0 + g(x_0)\}$$

$$= \left\{\frac{f(x_n)}{\prod\limits_{i \neq n}(x_n - x_i)} + \frac{f(x_{n-1})}{\prod\limits_{i \neq n-1}(x_{n-1} - x_i)} + \ldots. + \frac{f(x_0)}{\prod\limits_{i \neq 0}(x_0 - x_i)}\right\}$$

$$+ \left\{\frac{g(x_n)}{\prod\limits_{i \neq n}(x_n - x_i)} + \frac{g(x_{n-1})}{\prod\limits_{i \neq n-1}(x_{n-1} - x_i)} + \ldots. + \frac{g(x_0)}{\prod\limits_{i \neq 0}(x_0 - x_i)}\right\}$$

$$= f(x_0, x_1, \ldots x_n) + g(x_0, x_1 \ldots x_n).$$

Ex. 4-5. *Show that the divided differences of '$cf(x)$', where 'c' is constant, are 'c' times the divided differences of $f(x)$.*

Sol. Let $\phi(x) = cf(x)$

Then $$\phi(x_0, x_1, \ldots. x_n) = \frac{\phi(x_n)}{\prod\limits_{i \neq n}(x_n - x_i)} + \frac{\phi(x_{n-1})}{\prod\limits_{i \neq n-1}(x_{n-1} - x_i)} + \ldots. + \frac{\phi(x_0)}{\prod\limits_{i \neq 0}(x_0 - x_i)}$$

$$= c\left\{\frac{f(x_n)}{\prod\limits_{i\neq n}(x_n - x_i)} + \frac{f(x_{n-1})}{\prod\limits_{i\neq n-1}(x_{n-1} - x_i)} + \ldots + \frac{f(x_0)}{\prod\limits_{i\neq 0}(x_0 - x_i)}\right\}$$

$$= cf(x_0, x_1 \ldots x_n).$$

Ex. 4-6. *Show that nth order divided differences of x^n are constant.*

Sol. Let $f(x) = x^n$

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{x_1^n - x_0^n}{x_1 - x_0} = x_1^{n-1} + x_1^{n-2}x_0 + \ldots + x_0^{n-1}$$

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

$$= \frac{\left(x_2^{n-1} + x_2^{n-2}x_1 + \ldots + x_1^{n-1}\right) - \left(x_0^{n-1} + x_0^{n-2}x_1 + \ldots + x_1^{n-1}\right)}{x_2 - x_0}$$

$$= \frac{\left(x_2^{n-1} - x_0^{n-1}\right) + x_1\left(x_2^{n-2} - x_0^{n-2}\right) + \ldots + x_1^{n-2}(x_2 - x_0)}{x_2 - x_0}$$

$$= \left(x_2^{n-2} + x_2^{n-3}x_0 + \ldots + x_0^{n-2}\right)$$

$$+ x_1\left(x_2^{n-3} + x_2^{n-4}x_0 + \ldots + x_0^{n-3}\right) + \ldots + x_1^{n-2}$$

Thus if $f(x) = x^n$, $f(x_0, x_1)$ is a homogeneous function of degree $(n-1)$ *in* x_0, x_1; $f(x_0, x_1, x_2)$ is a homogeneous function of degree $(n-2)$ in x_0, x_1, x_2, and so on. Thus the operation of taking the divided difference lowers the degree by unity. Hence, finally, $f(x_0, x_1, \ldots x_n)$ will be a homogeneous function of degree $n - n = 0$ *i.e.*, a constant.

Ex. 4-7. *Prove that the third order divided difference with arguments a, b, c, d of the function $\frac{1}{x}$ is equal to $-\frac{1}{abcd}$.*

Sol.

x	$f(x) = \frac{1}{x}$	1st order	2nd order	3rd order
a	$\frac{1}{a}$	$-\frac{1}{ab}$		
b	$\frac{1}{b}$	$-\frac{1}{bc}$	$\frac{1}{abc}$	$-\frac{1}{abcd}$
c	$\frac{1}{c}$	$-\frac{1}{cd}$	$\frac{1}{bcd}$	
d	$\frac{1}{d}$			

4.2. Descending and Ascending Differences

Descending Differences. The first descending difference of $f(x)$ is defined by

$$\Delta f(x) = f(x+h) - f(x)$$

where h is the increment in x. The operator 'Δ' is called descending or forward difference operator. The second, third, etc., differences are defined by $\Delta\{\Delta f(x)\}$, $\Delta[\Delta\{\Delta f(x)\}]$ etc.

Operator E. The extensio

$Ef(x) = f(x$

Ascending Difference. Th

$\nabla f(x) = f(x$

The operator '∇' is called

third, etc., differences are defin

Central Differences. The

$\delta f(x) = f($

The operator 'δ' is called ce

are defined by $\delta\{\delta f(x)\}$, $\delta[\delta\{\delta$

Central Mean Operator.

$\mu = \frac{1}{2}$

Relations between Opera

(*ii*) $\nabla \equiv \frac{E-1}{E} \equiv \frac{\Delta}{E} \equiv$

(*iii*) $\Delta \equiv \frac{\nabla}{1-\nabla}$

(*iv*) $\delta \equiv E^{1/2} - E^{-1/2} \equiv \Delta E$

Relation between divided

$f(x_0, x_1 \ldots x_n) =$

Factorial Notation

$x^{(mh)} = x($

$x^{(-mh)} = (x$

Ex. 4-8. *Given $U_0 = 3$, U_1*

Sol. Difference table is

x	U(x)
0	3
1	12
2	81
3	200
4	100
5	8

$\therefore\ \Delta^5 U_0 = 755.$

Ex. 4-9. *Show that $E \equiv 1$*

Sol. By def. $\Delta f(x) = f($

where $1 f(x) = f(x)$

$$\ldots + \ldots + \frac{f(x_0)}{\prod\limits_{i \neq 0} (x_0 - x_i)}\Bigg\}$$

are constant.

$$+x_1^{n-2} x_0 + \ldots + x_0^{n-1}$$

$$x_0^{n-1} + x_0^{n-2} x_1 + \ldots + x_1^{n-1})$$

$$) + \ldots + x_1^{n-2} (x_2 - x_0)$$

$$+ \ldots\ldots + x_1^{n-2}$$

degree $(n-1)$ *in* x_0, x_1,; $f(x_0, x_1$, and so on. Thus the operation . Hence, finally, $f(x_0, x_1, \ldots x_n)$ onstant.

vith arguments a, b, c, d of the

er	d order

$$-\frac{1}{abcd}$$

e of $f(x)$ is defined by

ending or forward difference $\{\Delta f(x)\}$, $\Delta[\Delta\{\Delta f(x)\}]$ etc.

Operator E. The extension or shift operator 'E' is defined by

$$Ef(x) = f(x+h)$$

Ascending Difference. The first ascending difference of $f(x)$ is defined by

$$\nabla f(x) = f(x) - f(x-h)$$

The operator '∇' is called ascending or backward difference operator. The second, third, etc., differences are defined by $\nabla\{\nabla f(x)\}$, $\nabla[\nabla\{\nabla f(x)\}]$ etc.

Central Differences. The first central difference of $f(x)$ is defined by

$$\delta f(x) = f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right)$$

The operator 'δ' is called central difference operator. The second, third, etc., differences are defined by $\delta\{\delta f(x)\}$, $\delta[\delta\{\delta f(x)\}]$ etc.

Central Mean Operator. It is defined by

$$\mu = \frac{1}{2}\{E^{1/2} + E^{-1/2}\}$$

Relations between Operators. (*i*) $\Delta \equiv E - 1$

(*ii*) $\nabla \equiv \dfrac{E-1}{E} \equiv \dfrac{\Delta}{E} \equiv \dfrac{\Delta}{1+\Delta}$.

(*iii*) $\Delta \equiv \dfrac{\nabla}{1-\nabla}$

(*iv*) $\delta \equiv E^{1/2} - E^{-1/2} \equiv \Delta E^{-1/2} \equiv \nabla E^{1/2}$

Relation between divided differences and ordinary differences is :

$$f(x_0, x_1 \ldots\ldots x_n) = \frac{\Delta^n f(x_0)}{n!\,h^n}$$

Factorial Notation

$$x^{(mh)} = x(x-h)(x-2h) \ldots\ldots (x-\overline{m-1}\,h)$$

$$x^{(-mh)} = (x+h)^{-1}(x+2h)^{-1} \ldots\ldots (x+mh)^{-1}$$

Ex. 4-8. *Given* $U_0 = 3$, $U_1 = 12$, $U_2 = 81$, $U_3 = 200$, $U_4 = 100$ *and* $U_5 = 8$. *Find* $\Delta^5 U_0$.

Sol. Difference table is

x	$U(x)$	Δ	Δ^2	Δ^3	Δ^4	Δ^5
0	3					
		9				
1	12		60			
		69		– 10		
2	81		50		–259	
		119		–269		755
3	200		–219		496	
		–100		227		
4	100		8			
		–92				
5	8					

$\therefore\ \Delta^5 U_0 = 755.$

Ex. 4-9. *Show that* $E \equiv 1 + \Delta$.

Sol. By def.

$$\Delta f(x) = f(x+h) - f(x) = Ef(x) - f(x) = (E-1)f(x)$$

where $1 f(x) = f(x)$

$$\therefore \quad \Delta \equiv E-1 \quad \text{or} \quad E \equiv 1+\Delta$$

Ex. 4-10. *Show that $E \equiv e^{hD}$ where D denotes the derivative operator and deduce that $\Delta \equiv e^{hD}-1$.*

Sol. By def,
$$Ef(x) = f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \ldots$$
$$= \left(1+hD+\frac{h^2}{2!}D^2+\ldots\right)f(x)$$
$$= e^{hD}f(x)$$
$$E \equiv e^{hD}$$
$$1+\Delta \equiv e^{hD} \quad \text{or} \quad \Delta \equiv e^{hD}-1.$$

Ex. 4-11. *Show that*

(i) $$Dy = \frac{1}{h}\left(\Delta y - \frac{1}{2}\Delta^2 y + \frac{1}{3}\Delta^3 y - \frac{1}{4}\Delta^4 y + \ldots\right)$$

(ii) $$D^2 y = \frac{1}{h^2}\left(\nabla^2 y + \nabla^3 y + \frac{11}{12}\nabla^4 y + \ldots\right)$$

Sol. From Ex. 4-10, $e^{hD} \equiv 1+\Delta$

(i) $$\therefore \quad D \equiv \frac{1}{h}\log(1+\Delta)$$
$$\equiv \frac{1}{h}\left\{\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 \ldots\right\}$$
$$\therefore \quad Dy = \frac{1}{h}\left\{\Delta y - \frac{1}{2}\Delta^2 y + \frac{1}{3}\Delta^3 y \ldots\right\}$$

(ii) $$e^{hD} = 1 + \frac{\nabla}{1-\nabla} \equiv \frac{1}{1-\nabla}$$
$$\therefore \quad hD \equiv -\log(1-\nabla) \equiv \nabla + \frac{1}{2}\nabla^2 + \frac{1}{3}\nabla^3 + \frac{1}{4}\nabla^4 + \ldots$$
$$\therefore \quad h^2D^2 \equiv \left(\nabla + \frac{1}{2}\nabla^2 + \frac{1}{3}\nabla^3 + \frac{1}{4}\nabla^4 + \ldots\right)^2$$
$$\equiv \nabla^2 + \nabla^3 + \frac{11}{12}\nabla^4 + \ldots$$
$$\therefore \quad D^2 y = \frac{1}{h^2}\left(\nabla^2 y + \nabla^3 y + \frac{11}{12}\nabla^4 y + \ldots\right).$$

Ex. 4-12. *Show that*

(i) $\Delta\{f(x) \pm g(x)\} = \Delta f(x) \pm \Delta g(x)$
(ii) $Ef\{(x) \pm g(x)\} = Ef(x) \pm Eg(x)$
(iii) $\Delta\{cf(x)\} = c\Delta f(x)$
(iv) $E\{cf(x)\} = cEf(x)$
(v) $\Delta E \equiv E\Delta$
(vi) $\Delta^m \Delta^n \equiv \Delta^n \Delta^m \equiv \Delta^{m+n}$
(vii) $E^m E^n \equiv E^n E^m \equiv E^{n+m}$
(viii) $\Delta\{f.g\} = f(x+h)\,\Delta g(x) + g(x)\,\Delta f(x)$

(ix) $\Delta\left\{\frac{f(x)}{g(x)}\right\} = \frac{g(x)\Delta f(x)-}{g(x)g}$

Sol. (i) $\Delta\{f(x) \pm g(x)\} =$
$=$
$=$
(ii) $E\{f(x) \pm g(x)\} =$
(iii) $\Delta\{cf(x)\} =$
$=$
(iv) $E\{c.f(x)\} =$
(v) $\Delta Ef(x) =$
$=$
$\therefore \quad \Delta E \equiv$
(vi) $\Delta^m \Delta^n f(x) =$
$=$
$\therefore \quad \Delta^m \Delta^n \equiv$
Similarly $\Delta^n \Delta^m \equiv$
(vii) $E^m E^n f(x) =$
$=$
$\therefore \quad E^m E^n \equiv$
(viii) $\Delta\{f.g\} =$
$=$
$=$
(ix) $\Delta\left\{\frac{f(x)}{g(x)}\right\} =$
$=$
$=$

Ex. 4-13. *If (i) $f(E)$ is a poly*
$f(E)a^x =$
(ii) $f(\Delta)$ is a polynomia
$f(\Delta)a^x =$

Sol. (i) Let $f(E) \equiv$
$\therefore \quad f(E)\,a^x =$
$=$
$=$
$=$
(ii) Let $f(\Delta) \equiv$
$\therefore \quad f(\Delta)\,a^x =$
$=$
$=$
$=$

Ex. 4-14. *Show that*

$e^{-x} =$

tive operator and deduce that

$'(x) + ...$

$+\frac{1}{4}\nabla^4 + ...$

(*ix*) $\Delta\left\{\dfrac{f(x)}{g(x)}\right\} = \dfrac{g(x)\Delta f(x) - f(x)\Delta g(x)}{g(x)g(x+h)}$

Sol. (*i*)

$$\begin{aligned} \Delta\{f(x) \pm g(x)\} &= \{f(x+h) \pm g(x+h)\} - \{f(x) \pm g(x)\} \\ &= \{f(x+h) - f(x)\} \pm \{g(x+h) - g(x)\} \\ &= \Delta f(x) \pm \Delta g(x) \end{aligned}$$

(*ii*) $E\{f(x) \pm g(x)\} = f(x+h) \pm g(x+h) = Ef(x) \pm Eg(x)$

(*iii*)

$$\begin{aligned} \Delta\{cf(x)\} &= cf(x+h) - cf(x) \\ &= c\{f(x+h) - f(x)\} = c\Delta f(x) \end{aligned}$$

(*iv*) $E\{c.f(x)\} = cf(x+h) = cEf(x)$

(*v*)

$$\begin{aligned} \Delta Ef(x) &= \Delta\{f(x+h) = f(x+2h) - f(x+h) \\ &= E\{f(x+h) - f(x)\} = E\Delta f(x) \\ \therefore\quad \Delta E &\equiv E\Delta \end{aligned}$$

(*vi*)

$$\begin{aligned} \Delta^m\Delta^n f(x) &= (\Delta\Delta\ ...\ m \text{ times})\ (\Delta\Delta ...\ n \text{ times})\, f(x) \\ &= (\Delta\Delta ...\ (m+n) \text{ times})\, f(x) = \Delta^{m+n} f(x) \\ \therefore\quad \Delta^m\Delta^n &\equiv \Delta^{m+n} \end{aligned}$$

Similarly $\Delta^n\Delta^m \equiv \Delta^{m+n}$

(*vii*)

$$\begin{aligned} E^mE^n f(x) &= (EE...m \text{ times})\ (EE...n \text{ times}) \\ &= \{EE...(m+n) \text{ times}\}\, f(x) = E^{m+n} f(x) \\ \therefore\quad E^mE^n &\equiv E^{m+n} \end{aligned}$$

(*viii*)

$$\begin{aligned} \Delta\{f\,.\,g\} &= f(x+h).\, g(x+h) - f(x).\, g(x) \\ &= f(x+h)\ \{g(x+h) - g(x)\} + g(x)\ \{f(x+h) - f(x)\} \\ &= f(x+h)\ \Delta g(x) + g(x)\ \Delta f(x) \end{aligned}$$

(*ix*)

$$\begin{aligned} \Delta\left\{\frac{f(x)}{g(x)}\right\} &= \frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)} = \frac{g(x)f(x+h) - f(x)g(x+h)}{g(x)g(x+h)} \\ &= \frac{g(x)\{f(x+h) - f(x)\} - f(x)\{g(x+h) - g(x)\}}{g(x)g(x+h)} \\ &= \frac{g(x)\Delta f(x) - f(x)\Delta g(x)}{g(x)g(x+h)}. \end{aligned}$$

Ex. 4-13. *If* (*i*) $f(E)$ *is a polynomial in E, show that*

$$f(E)a^x = a^x f(a^h)$$

(*ii*) $f(\Delta)$ *is a polynomial in* Δ, *show that*

$$f(\Delta)a^x = a^x f(a^h - 1).$$

Sol. (*i*) Let

$$\begin{aligned} f(E) &\equiv p_0 + p_1 E + p_2 E^2 + ...\, p_n E^n \\ \therefore\quad f(E)\, a^x &= (p_0 + p_1 E + p_2 E^2 + ... + p_n E^n)\, a^x \\ &= p_0\, a^x + p_1 a^{x+h} + p_2\, a^{x+2h}\ ....\, p_n\, a^{x+nh} \\ &= a^x\ \{p_0 + p_1\, a^h + p_2 a^{2h} + ...\, p_n\, a^{nh}\} \\ &= a^x f(a^h) \end{aligned}$$

(*ii*) Let

$$\begin{aligned} f(\Delta) &\equiv p_0 + p_1\, \Delta^2 + ... + p_n\, \Delta^n \\ \therefore\quad f(\Delta)\, a^x &= (p_0 + p_1\, \Delta + ... + p_n\, \Delta^n)\, a^x \\ &= \{p_0\, a^x + p_1\, (E-1)\, a^x + p_2\, (E-1)^2\, a^x + ... + p_n\, (E-1)^n\, a^x\} \\ &= a^x\ \{p_0 + p_1\, (a^h - 1) + p_2\, (a^h - 1)^2 + ... + p_n\, (a^h - 1)^n] \\ &= a^x f(a^h - 1). \end{aligned}$$

Ex. 4-14. *Show that*

$$e^{-x} = \left(\frac{\Delta^2}{E}\right) e^{-x}\, .\, \frac{Ee^{-x}}{\Delta^2\, e^{-x}}\mathrm{s}$$

Sol. R.H.S. $= \left(\dfrac{\Delta^2}{E}\, e^{-x}\right) \dfrac{Ee^{-x}}{\Delta^2 e^{-x}}$

$$= \Delta^2 (E^{-1} e^{-x}) \cdot \frac{(Ee^{-x})}{\Delta^2 e^{-x}} = \Delta^2 (e^{-x+h}) \cdot \frac{e^{-x-h}}{\Delta^2 e^{-x}}$$

$$= (e^h \Delta^2 e^{-x}) \cdot \frac{e^{-x} . e^{-h}}{\Delta^2 e^{-x}} = e^{-x} = \text{L.H.S.}$$

Ex. 4-15. *Show that* $\Delta (\tan^{-1} x) = \tan^{-1}\left\{\dfrac{h}{1+xh+x^2}\right\}$*, where h is the interval of differencing.*

Sol. L.H.S. $= \tan^{-1}(x+h) - \tan^{-1} x$

$$= \tan^{-1}\left\{\frac{x+h-x}{1+x(x+h)}\right\} = \tan^{-1}\left\{\frac{h}{1+xh+x^2}\right\}.$$

Ex. 4-16. *Explain the difference between* $\left(\dfrac{\Delta^2}{E}\right) U_x$ *and* $\left(\dfrac{\Delta^2 U_x}{EU_x}\right)$ *and find the values of these functions when* $U_x = x^3$.

Sol. $\left(\dfrac{\Delta^2}{E}\right) U_x$ = Result of operating $\Delta^2 E^{-1}$ on U_x

and $\left(\dfrac{\Delta^2 U_x}{EU_x}\right)$ = Ratio of the results of operating Δ^2 and E on U_x.

$$\left(\frac{\Delta^2}{E}\right) x^3 = \Delta^2 E^{-1}(x^3) = \Delta^2 (x-h)^3 = \Delta\{x^3 - (x-h)^3\}$$

$$= (x+h)^3 - 2x^3 + (x-h)^3 = 6xh^2$$

$$\frac{\Delta^2 x^3}{Ex^3} = \frac{(x+2h)^3 - 2(x+h)^3 + (x^3)}{(x+h)^3} = \frac{6xh^2 + 6h^3}{(x+h)^3} = \frac{6h^2}{(x+h)^2}$$

Ex. 4-17. *If* $f(x) = e^{ax}$*, show that* $f(x)$ *and its leading differences are in G.P.*

Sol.

$$f(x) = e^{ax},\ \Delta f(x) = e^{a(x+h)} - e^{ax} = e^{ax}\{e^{ah} - 1\}$$

$$\Delta^2 f(x) = \Delta\{e^{ax}(e^{ah} - 1)\} = (e^{ah} - 1)\,\Delta e^{ax}$$

$$= (e^{ah} - 1)^2 e^{ax}.$$

Similarly, $\Delta^3 f(x) = (e^{ah} - 1)^3 e^{ax}$ and so on.

Evidently $f(x)$, $\Delta f(x)$, $\Delta^2 f(x)$ .... are in G.P.

Ex. 4-18. *Show that* (*i*) $\Delta^n \equiv E^n - {}^nc_1 E^{n-1} + {}^nc_2 E^{n-2} \ldots + (-1)^n$

(*ii*) $E^n \equiv \Delta^n + {}^nc_1 \Delta^{n-1} + {}^nc_2 \Delta^{n-2} + \ldots + 1$.

Sol. (*i*) By def., $\Delta f(x) = f(x+h) - f(x) = (E-1) f(x)$

$\therefore$ $\Delta \equiv E - 1$

$$\Delta^2 f(x) = f(x+2h) - 2f(x+h) + f(x) = (E^2 - 2E + 1) f(x)$$

$\therefore$ $\Delta^2 \equiv E^2 - 2E + 1$

Let, in general, $\Delta^r \equiv E^r - {}^rc_1 E^{r-1} + \ldots + (-1)^k\, {}^rc_k E^{r-k} + \ldots + (-1)^r$

Then $\Delta^{r+1} \equiv \Delta\Delta^r \equiv (E-1)(E^r - {}^rc_1 E^{r-1} + \ldots + (-1)^k\, {}^rc_k E^{r-k} + \ldots + (-1)^r)$

$$\equiv E^{r+1} - ({}^rc_1 + 1) E^r + ({}^rc_2 + {}^rc_1) E^{r-1} \ldots +$$

$$(-1)^k ({}^rc_k + {}^rc_{k-1}) E^{r+1-k} \ldots - (-1)^r$$

$\equiv E^{r+1} - {}^{r+1}c_1 + E$
.... $+ (-1)^{r+1}$.
$\therefore$ If the result holds for $n = r$, it a
hold for $n = 2$, hence it holds for $n =$
(*ii*) left as an exercise.
Ex. 4-19. *If* $y_x = \sin x$*, show tha*
Sol. $\Delta^2 y_x$ = (E
= sin
= 2 s
= kE
where $k = 2(\cos h - 1)$.
Ex. 4-20. *Show that* $\Delta^n x^n$ =
polynomial of degree 'n' is constant.
Sol. $\Delta^n x^n = \Delta^n$
$= \Delta^n$
$= \Delta$
$= \Delta^n$
+ .
$= \Delta^n$
where a_3, a_4 etc., depend upon n and
$\Delta^n x^n = \Delta\{$
where b_0 is also constant depending
$\therefore$ $\Delta^n x^n = n\{$
Now consider a polynomial
$f(x) = A_0$
Then $\Delta^n f(x) = A_0$
$= A_0$
Ex. 4-21. *Show that* $\Delta^{10}(1 - ax$
Sol. Let $\phi(x) = (1$
= ab
$\Delta^{10} \phi(x) = 10$
Ex. 4-22. *If* $x_0, x_1, \ldots x_n$ *be equ*
$f(x_0, x_1, \ldots x_n) =$
Sol. Let $x_{i+1} - x_i = h$
By def., $f(x_0, x_1) =$
$f(x_0, x_1, x_2) =$
=
Let $f(x_0, x_1, \ldots x_m) =$
Then $f(x_0, x_1, \ldots x_{m+1}) =$

$$\equiv E^{r+1} - {}^{r+1}c_1 + E^{\overline{r+1}-1} + {}^{r+1}c_2\, E^{\overline{r+1}-2} - \ldots + (-1)^k\, {}^{r+1}c_k\, E^{\overline{r+1}-k}\, r \;\ldots + (-1)^{r+1}.$$

$\therefore$ If the result holds for $n = r$, it also holds for $n = r + 1$. But the result has been seen to hold for $n = 2$, hence it holds for $n = 2 + 1 = 3$ and hence for $n = 3 + 1 = 4$ and so on.

(*ii*) left as an exercise.

Ex. 4-19. *If $y_x = \sin x$, show that $\Delta^2 y_x = kEy_x$, where k is some constant.*

Sol.
$$\begin{aligned}\Delta^2 y_x &= (E-1)^2 \sin x = (E^2 - 2E + 1)\sin x\\ &= \sin(x+2h) - 2\sin(x+h) + \sin x\\ &= 2\sin(x+h)\{\cos h - 1\} = k \sin(x+h)\\ &= kE(\sin x)\end{aligned}$$

where $\quad k = 2(\cos h - 1)$.

Ex. 4-20. *Show that $\Delta^n x^n = n!h^n$ and deduce that nth descending difference of a polynomial of degree 'n' is constant.*

Sol.
$$\begin{aligned}\Delta^n x^n &= \Delta^{n-1}\{\Delta x^n\} = \Delta^{n-1}\{(x+h)^n - x^n\}\\ &= \Delta^{n-1}\{nhx^{n-1} + {}^nc_2\, h^2 x^{n-2} + \ldots + h^n\}\\ &= \Delta^{n-2}\{nh\Delta x^{n-1} + {}^nc_2\, h^2\, \Delta x^{n-2} + \ldots + nh^{n-1}\,\Delta x\}\\ &= \Delta^{n-2}[nh\{(x+h)^{n-1} - x^{n-1}\} + {}^nc_2\, h^2\, \{(x+h)^{n-2} - x^{n-2}\}\\ &\quad + \ldots + nh^{n-1}(x + h - x)]\\ &= \Delta^{n-2}[n(n-1)h^2 x^{n-2} + a_3 x^{n-3} + a_4 x^{n-4} + \ldots + a_n]\end{aligned}$$

where a_3, a_4 etc., depend upon n and h. Proceeding likewise finally,

$$\Delta^n x^n = \Delta\{n(n-1)\ldots 2.h^{n-1}x + b_0\}$$

where b_0 is also constant depending upon n and h.

$$\therefore \quad \Delta^n x^n = n\{n-1)\ldots 2.h^{n-1}(x + h - x) = n!h^n$$

Now consider a polynomial

$$f(x) = A_0 x^n + A_1 x^{n-1} + \ldots + A_n$$

Then
$$\begin{aligned}\Delta^n f(x) &= A_0\,\Delta^n x^n + A_1\,\Delta^n x^{n-1} + \ldots\, \Delta^n A_n\\ &= A_0\,\Delta^n x^n = A_0\, n!h^n = \text{constant}.\end{aligned}$$

Ex. 4-21. *Show that* $\Delta^{10}(1-ax)(1-bx^2)(1-cx^3)(1-dx^4) = abcd\,10!$

Sol. Let
$$\begin{aligned}\phi(x) &= (1-ax)(1-bx^2)(1-cx^3)(1-dx^4)\\ &= abcd\,x^{10} + \text{terms containing lower powers of } x\\ \Delta^{10}\phi(x) &= 10!\,abcd \text{ (taking } h = 1).\end{aligned}$$

Ex. 4-22. *If $x_0, x_1, \ldots x_n$ be equally spaced, show that*

$$f(x_0, x_1, \ldots x_n) = \frac{\Delta^n f(x_0)}{n!h^n}.$$

Sol. Let
$$x_{i+1} - x_i = h\ (i = 0,\ 1, \ldots n-1)$$

By def.,
$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0+h) - f(x_0)}{h} = \frac{\Delta f(x_0)}{h}$$

$$\begin{aligned}f(x_0, x_1, x_2) &= \frac{f(x_1, x_2) - (x_0, x_1)}{x_2 - x_0} = \frac{1}{2h}\left\{\frac{\Delta f(x_1)}{h} - \frac{\Delta f(x_0)}{h}\right\}\\ &= \frac{1}{2h^2}\,\Delta\{f(x_0+h) - f(x_0)\} = \frac{\Delta^2 f(x_0)}{2!h^2}\end{aligned}$$

Let
$$f(x_0, x_1, \ldots x_m) = \frac{\Delta^m f(x_0)}{m!h^m}$$

Then
$$f(x_0, x_1, \ldots x_{m+1}) = \frac{f(x_1, \ldots x_{m+1}) - f(x_0, x_1, \ldots x_m)}{x_{m+1} - x_0}$$

$$= \frac{1}{(m+1)h}\left\{\frac{\Delta^m f(x_1)}{m!h^m} - \frac{\Delta^m f(x_0)}{m!h^m}\right\}$$

$$= \frac{1}{(m+1)!h^{m+1}}\,\Delta^m\,\{f(x_0+h) - f(x_0)\}$$

$$= \frac{\Delta^{m+1} f(x_0)}{(m+1)!h^{m+1}}.$$

∴ By induction,

$$f(x_0, x_1 \ldots\ldots x_n) = \frac{\Delta^n f(x_0)}{n!h^n} \quad \text{for all } n \ge 1.$$

Ex. 4-23. *Show that* (*i*) $\nabla \equiv \dfrac{E-1}{E}$, (*ii*) $\nabla E \equiv \Delta \equiv E\nabla$.

Sol. (*i*) By def,

$$\nabla f(x) \equiv f(x) - f(x-h)$$
$$\equiv f(x) - E^{-1} f(x) \equiv (1 - E^{-1}) f(x)$$

$$\therefore \quad \nabla \equiv 1 - \frac{1}{E} \equiv \frac{E-1}{E}$$

(*ii*)
$$\nabla E \equiv (1 - E^{-1})\,E \equiv E - 1 = \Delta$$
$$E\nabla \equiv E(1 - E^{-1}) \equiv E - 1 = \Delta.$$

Ex. 4-24. *Show that* (*i*) $(1+\Delta)(1-\nabla) \equiv 1$

(*ii*) $\Delta\nabla \equiv \Delta - \nabla$

Sol. (*i*) $(1+\Delta)(1-\nabla) \equiv E(1-\nabla) \equiv E - E\nabla \equiv E - \Delta \equiv 1$

(*ii*) $\Delta\nabla \equiv (E-1)\,\nabla \equiv E\nabla - \nabla \equiv \Delta - \nabla$.

Ex. 4-25. *Show that* (*i*) $\Delta^n x^{(mh)} = (mh)^{(nh)}\, x^{(\overline{m-n}\,h)}$

(*ii*) $\Delta\,(x^n)^{(-mh)} = (-mh)^{(nh)}\, x^{(-m+nh)}$

Sol.
$$\Delta x^{(mh)} = (x+h)^{(mh)} - x^{(mh)}$$
$$= (x+h)\,(x)\,(x-h) \ldots. (x+h-\overline{m-1}h) - x(x-h) \ldots (x - \overline{m-1}h)$$
$$= x(x-h) \ldots. (x - \overline{m-2}h)\ \{(x+h) - (x - \overline{m-1}h)\}$$
$$= (mh)\,x^{(\overline{m-1}h)}$$

$$\therefore \quad \Delta^2 x^{(mh)} = (mh)(\overline{m-1}h)\,x^{(\overline{m-2}h)} = (mh)^{(2h)}\, x^{\overline{m-2h}}$$

Proceeding likewise finally

$$\Delta^n x^{(mh)} = [mh]^{(nh)}\, x^{(\overline{m-nh})}$$

(*ii*) Left as an exercise.

Ex. 4-26. *Show that* (*i*) $\Delta ab^{cx} = (b^c - 1)\, ab^{cx}$

(*ii*) $\Delta x^{(r)} = rx^{(r-1)}$

Sol. (*i*) $\Delta ab^{cx} = ab^{c(x+1)} - ab^{cx} = ab^{cx}\,(b^c - 1)$

(*ii*) See Ex. 4-25.

Ex. 4-27. *Show that*

(*i*) $n(n-1) + (n-1)(n-2) + \ldots. + 2.1 = \dfrac{1}{3}(n+1)\,n\,(n-1)$

(*ii*) $n(n-1)(n-2) + (n-1)(n-2)(n-3) + \ldots. + 3.2.1. = \dfrac{1}{4}(n+1)\,n\,(n-1)(n-2)$

Sol. (*i*) Let
$$S = n(n-1) + (n-1)(n-2) + \ldots.. + 2.1$$
$$= n^{(2)} + (n-1)^{(2)} + \ldots.. + 2^{(2)}$$

Now $\Delta n^{(3)} = 3n^{(2}$

∴ $n^{(2)} = \frac{1}{3}\{$

Changing n to $n-1, n-2, \ldots.. 3$

$(n-1)^{(2)} = \frac{1}{3}\{$

$(n-2)^{(2)} = \frac{1}{3}\{$

....

$3^{(2)} = \frac{1}{3}$

Adding $S = \frac{1}{3}$

$= \frac{1}{3}$

$= \frac{1}{3}($

(*ii*) Left as an exercise.

Ex. 4-28. *Express* $x^3 - 3x + 1$ *in* ... *difference.*

Sol. Let $\phi(x) = x^3$

Then $\Delta\phi(x) = a_1$

$\Delta^2\phi(x) = 2^{(1}$

$= 2^{(2}$

and $\Delta^3\phi(x) = 3^{(2}$

∴ $a_3 = \frac{1}{6}$

$a_2 = \frac{1}{2}$

$a_1 = \Delta$

$a_0 = \phi($

∴ $\phi(x) = \phi($

Difference table of $\phi(x)$ is given

x	$\phi(x)$
0	1
1	−1
2	3
3	19

∴ $\phi(x) = 1$

∴ $\Delta^2\phi(x) = 6$

Ex. 4.29. *A third degree polyn*... (2, 1), *and* (3, − 2). *Find the value a*

Now $\quad \Delta n^{(3)} = 3n^{(2)}$

$$\therefore \quad n^{(2)} = \frac{1}{3}\{(n+1)^{(3)} - n^{(3)}\}$$

Changing n to $n-1, n-2, \ldots, 3$

$$(n-1)^{(2)} = \frac{1}{3}\{n^{(3)} - (n-1)^{(3)}\}$$

$$(n-2)^{(2)} = \frac{1}{3}\{(n-1)^{(3)} - (n-2)^{(3)}\}$$

....

$$3^{(2)} = \frac{1}{3}\{4^{(3)} - 3^{(3)}\}$$

Adding

$$S = \frac{1}{3}\{(n+1)^{(3)} - 3^{(3)}\} + 2^{(2)}$$

$$= \frac{1}{3}\{(n+1)\,n\,(n-1) - 3.2.1\} + 2.1$$

$$= \frac{1}{3}(n+1)\,n(n-1)$$

(*ii*) Left as an exercise.

Ex. 4-28. *Express* $x^3 - 3x + 1$ *in the factorial notation and use it to obtain its second difference.*

Sol. Let $\quad \phi(x) = x^3 - 3x + 1 \equiv a_0 + a_1x^{(1)} + a_2x^{(2)} + a_3x^{(3)}$

Then

$$\Delta\phi(x) = a_1 + 2^{(1)}\,a_2\,x^{(1)} + 3^{(1)}\,a_3x^{(2)}$$

$$\Delta^2\phi(x) = 2^{(1)}\,a_2 + 3^{(1)}\,.\,2^{(1)}\,a_3\,x^{(1)}$$

$$= 2^{(2)}\,a_2 + 3^{(2)}\,a_3x^{(1)}$$

and

$$\Delta^3\,\phi(x) = 3^{(2)}\,a_3 = 3^{(3)}\,a_3 = 6a_3$$

$$\therefore \quad a_3 = \frac{1}{6}\,\Delta^3\,\phi(0)$$

$$a_2 = \frac{1}{2}\,\Delta^2\,\phi(0)$$

$$a_1 = \Delta\,\phi(0)$$

$$a_0 = \phi(0)$$

$$\therefore \quad \phi(x) = \phi(0) + \Delta\phi\,(0)\,x^{(1)} + \frac{1}{2}\,\Delta^2\,\phi(0)\,x^{(2)} + \frac{1}{6}\,\Delta^3\,\phi(0)x^{(3)}$$

Difference table of $\phi(x)$ is given below :

x	$\phi(x)$	Δ	Δ^2	Δ^3
0	1			
		–2		
1	–1		6	
		4		6
2	3		12	
		16		
3	19			

$$\therefore \quad \phi(x) = 1 - 2x^{(1)} + 3x^{(2)} + x^{(3)}$$

$$\therefore \quad \Delta^2\,\phi(x) = 6 + 6x^{(1)} = 6 + 6x.$$

Ex. 4.29. *A third degree polynomial* $f(x)$ *is passed through the points* (0, –1), (1, 1), (2, 1), *and* (3, – 2). *Find the value at* $x = 1{\cdot}2$.

Sol. Difference table of $f(x)$ is

x	$f(x)$	Δ	Δ^2	Δ^3
0	-1			
		2		
1	1		-2	
		0		-1
2	1		-3	
		-3		
3	-2			

$$\therefore \quad f(x) = f(0) + \Delta f(0)\, x^{(1)} + \frac{1}{2}\Delta^2 f(0)\, x^{(2)} + \frac{1}{6}\Delta^3 f(0)\, x^{(3)}$$

$$= -1 + 2x^{(1)} - x^{(2)} - \frac{1}{6}x^{(3)}$$

$$= -1 + 2x - x(x-1) - \frac{1}{6}x(x-1)(x-2)$$

$$= x\left[(x-1)\left\{-\frac{1}{6}(x-2)-1\right\}+2\right]-1$$

Let
$$g_0(x) = -\frac{1}{6}, \qquad g_1(x) = g_0(x)(x-2)-1,$$
$$g_2(x) = (x-1)g_1(x)+2, \qquad g_3(x) = xg_2(x)-1 = f(x).$$

$$\therefore \quad g_0(1\cdot2) = -\frac{1}{6},\ g_1(1\cdot2) = \left(-\frac{1}{6}\right)(-0\cdot8)-1 = \frac{-2\cdot6}{3}$$

$$g_2(1\cdot2) = (0\cdot2)\left(-\frac{2\cdot6}{3}\right)+2 = \frac{5\cdot48}{3}$$

$$g_3(1\cdot2) = (1\cdot2)\left(\frac{5\cdot48}{3}\right) - 1 = 1\cdot192 = f(x)$$

The calculations are best carried out using the following computational scheme which is clearly related to synthetic division.

$$c_0 = g_0 \rightarrow (x-2) \rightarrow (x-2)\,g_0$$
$$\frac{c_1}{g_1}$$
$$(x-1)g_1 \leftarrow (x-1) \leftarrow$$
$$\frac{c_2}{g_2} \rightarrow x \rightarrow xg_2$$
$$\frac{c_3}{g_3 = f(x)}$$

For the above question.

$$-\frac{1}{6} \rightarrow (1\cdot2-2) \rightarrow \frac{0\cdot4}{3}$$
$$\frac{-1}{\;}$$
$$-\frac{0\cdot52}{3} \leftarrow (1\cdot2-1) \leftarrow \frac{2\cdot6}{3}$$
$$\frac{2}{\;}$$
$$\frac{5\cdot48}{3} \rightarrow (1\cdot2) \rightarrow 2\cdot192$$
$$\frac{-1}{\;}$$
$$\underline{1\cdot192 = f(x)}$$

Ex. 4-30. *Show that*

$$y_0 + y_1\frac{x}{1!} + y_2\frac{x^2}{2!} + y_3$$

Sol. L.H.S. =

=

=

=

Ex. 4-31. *Show that*

$$xy_1 + x^2y_2 + x^3y_3 + \ldots =$$

Sol. L.H.S. =

=

=

=

=

Ex. 4-32. *Show that*

$$y_x = y_{x-1} + \Delta y_{x-2} + \Delta^2$$

Sol. R.H.S. =

=

=

=

=

Ex. 4-33. *Show that*

$$\Delta x^m - \frac{1}{2}\Delta^2 x^m + \frac{1.3}{2.4}$$

Ex. 4-30. *Show that*

$$y_0 + y_1 \frac{x}{1!} + y_2 \frac{x^2}{2!} + y_3 \frac{x^3}{3!} + \ldots\ldots = e^x \left\{ y_0 + x\Delta y_0 + \frac{x^2}{2!} \Delta^2 y_0 + \ldots \right\}.$$

Sol.

$$\text{L.H.S.} = \left\{ 1 + \frac{x}{1!} E + \frac{x^2}{2!} E^2 + \ldots\right\} y_0$$

$$= e^{xE} y_0 = e^{x(1+\Delta)} y_0$$

$$= e^x . e^{x\Delta} y_0 = e^x \left\{ 1 + x\Delta + \frac{x^2 \Delta^2}{2!} + \ldots \right\} y_0$$

$$= e^x \left\{ y_0 + x\Delta y_0 + \frac{x^2}{2!} \Delta^2 y_0 + \ldots \right\} = \text{R.H.S.}$$

Ex. 4-31. *Show that*

$$xy_1 + x^2 y_2 + x^3 y_3 + \ldots = \frac{x}{1-x} y_1 + \frac{x^2}{(1-x)^2} \Delta y_1 + \frac{x^3}{(1-x)^3} \Delta^2 y_1 + \ldots\ldots$$

Sol.

$$\text{L.H.S.} = (xE + x^2E^2 + x^3E^3 + \ldots) y_0$$

$$= \frac{xE}{1-xE} y_0 = \frac{xE}{1-x-x\Delta} y_0$$

$$= \frac{xE}{1-x} \left[1 - \frac{x\Delta}{1-x} \right]^{-1} y_0$$

$$= \frac{xE}{1-x} \left\{ 1 + \frac{x}{1-x} \Delta + \frac{x^2}{(1-x)^2} \Delta^2 + \ldots \right\} y_0$$

$$= \frac{x}{1-x} y_1 + \frac{x^2}{(1-x)^2} \Delta y_1 + \frac{x^3}{(1-x)^3} \Delta^2 y_1 + \ldots\ldots$$

Ex. 4-32. *Show that*

$$y_x = y_{x-1} + \Delta y_{x-2} + \Delta^2 y_{x-3} + \ldots + \Delta^{n-1} y_{x-n} + \Delta^n y_{x-n}.$$

Sol.

$$\text{R.H.S.} = E^{-1} \left\{ 1 + \frac{\Delta}{E} + \frac{\Delta^2}{E^2} + \ldots + \frac{\Delta^{n-1}}{E^{n-1}} \right\} y_x + \Delta^n y_{x-n}$$

$$= E^{-1} \{1 + \nabla + \nabla^2 + \ldots + \nabla^{n-1}\} y_x + \Delta^n y_{x-n}$$

$$= E^{-1} \frac{(1-\nabla^n)}{1-\nabla} y_x + \Delta^n y_{x-n}$$

$$= \frac{(1-\nabla^n)}{E - E\nabla} y_x + \Delta^n y_{x-n} = \frac{1-\nabla^n}{E-\Delta} y_x + \frac{\Delta^n}{E^n} y_x$$

$$= (1-\nabla^n) y_x + \nabla^n y_x = y_x.$$

Ex. 4-33. *Show that*

$$\Delta x^m - \frac{1}{2} \Delta^2 x^m + \frac{1.3}{2.4} \Delta^3 x^m - \frac{1.3.5}{2.4.6} \Delta^4 x^m + \ldots + m \text{ terms}$$

$$= \Delta E^{-1/2} x^m = \left(x+\frac{1}{2}\right)^m - \left(x-\frac{1}{2}\right)^m$$

where $h = 1$

Sol. Since $\Delta^r x^m = 0$ for $r > m$,

$$\text{L.H.S.} = \Delta x^m - \frac{1}{2}\Delta^2 x^m + \frac{1.3}{2.4}\Delta^3 x^m - \frac{1.3.5}{2.4.6}\Delta^4 x^m + \text{.... up to } \infty$$

$$= \left(\Delta - \frac{1}{2}\Delta^2 + \frac{1.3}{2.4}\Delta^3 \ldots\ldots\right) x^m$$

$$\Delta(1+\Delta)^{-1/2} x^m = \Delta E^{-1/2} x^m$$

$$= \Delta\left(x-\frac{1}{2}\right)^m = \left(x+\frac{1}{2}\right)^m - \left(x-\frac{1}{2}\right)^m.$$

Ex. 4-34. *Show that*

$$y_{x+n} = y_n + {}^x c_1 \Delta y_{n-1} + {}^{x+1}c_2 \Delta^2 y_{n-2} + {}^{x+2}c_3 \Delta^3 y_{n-3} + \ldots.$$

Sol.

$$\text{R.H.S.} = \left(1 + {}^x c_1 \frac{\Delta}{E} + {}^{x+1}c_2 \frac{\Delta^2}{E^2} + {}^{x+2}c_3 \frac{\Delta^3}{E^3} + \ldots\ldots\right) y_n$$

$$= \left(1 + x\frac{\Delta}{E} + \frac{(x+1)}{2!} x \frac{\Delta^2}{E^2} + \frac{(x+2)(x+1)x}{3!}\frac{\Delta^3}{E^3} + \ldots\ldots\right) y_n$$

$$= \left(1 - \frac{\Delta}{E}\right)^{-x} y_n = E^x y_n = y_{x+n}.$$

Ex. 4-35. *Show that* $y_x - y_{x+1} + y_{x+2} - y_{x+3} + \ldots\ldots$

$$= \frac{1}{2}\left[y_{x-\frac{1}{2}} - \frac{1}{8}\Delta^2 y_{x-\frac{3}{2}} + \frac{1.3}{2!}\left(\frac{1}{8}\right)^2 \Delta^4 y_{x-\frac{5}{2}} - \frac{1.3.5}{3!}\left(\frac{1}{8}\right)^3 \Delta^6 y_{x-\frac{7}{2}} + \ldots\ldots \right]$$

Sol.

$$\text{R.H.S.} = \frac{1}{2} E^{-\frac{1}{2}} \left\{ 1 - \frac{1}{8}\frac{\Delta^2}{E} + \frac{1.3}{2!}\left(\frac{1}{8}\right)^2 \frac{\Delta^4}{E^2} - \frac{1.3.5}{3!}\left(\frac{1}{8}\right)^3 \frac{\Delta^6}{E^3} + \ldots \right\} y_x$$

$$= \frac{1}{2} E^{-1/2} \left(1 + \frac{1}{4}\frac{\Delta^2}{E}\right)^{-1/2} y_x$$

$$= \frac{1}{2}\left(E + \frac{1}{4}\Delta^2\right)^{-1/2} y_x = \frac{1}{2}\left\{E + \frac{1}{4}(E-1)^2\right\}^{-1/2} y_x$$

$$= \frac{1}{2}\left\{\frac{(E+1)^2}{4}\right\}^{-1/2} y_x = (E+1)^{-1} y_x$$

$$= (1 - E + E^2 - E^3 + \ldots\ldots) y_x$$

$$= y_x - y_{x+1} + y_{x+2} - y_{x+3} + \ldots\ldots = \text{L.H.S.}$$

Ex. 4-36. *Show that* $y_x - \frac{1}{8}\Delta^2 y_{x-1} + \frac{1.3}{8.16}\Delta^4 y_{x-2} - \frac{1.3.5}{8.16.24}\Delta^6 y_{x-3} + \ldots$

Sol. L.H.S. =

Ex. 4-37. *Show that* $y_0 + {}^x c_1$

Sol. R.H.S. =

Ex. 4-38. *Show that*

$\Delta^n y_{x-n} = y_x - {}^n c_1 y_{x-1}$

Sol. R.H.S. =

Ex. 4-39. *Show that*

$U_0 + U_1 + \ldots + U_n =$

Sol. L.H.S.

$\left(x-\frac{1}{2}\right)^m$

$^m - \frac{1.3.5}{2.4.6}\ \Delta^4 x^m + \ldots$ up to ∞

^{m}

^{m}

$-\left(x-\frac{1}{2}\right)^m.$

$^3 y_{n-3} + \ldots$

$^{x+2}c_3\ \frac{\Delta^3}{E^3} + \ldots \Big) y_n$

$\frac{(x+2)(x+1)x}{3!}\ \frac{\Delta^3}{E^3} + \ldots \Big) y_n$

$+ n.$

$\frac{.3}{2!}\left(\frac{1}{8}\right)^2 \Delta^4 y_{x-\frac{5}{2}}$

$\ldots \Big]$

$\frac{1.3.5}{3!}\left(\frac{1}{8}\right)^3 \frac{\Delta^6}{E^3} + \ldots \Big\} y_x$

$-1)^2 \Big\}^{-1/2} y_x$

$- \frac{1.3.5}{8.16.24}\ \Delta^6 y_{x-3} + \ldots$

$$= y_{x+\frac{1}{2}} - \frac{1}{2}\Delta y_{x+\frac{1}{2}} + \frac{1}{4}\Delta^2 y_{x+\frac{1}{2}} - \frac{1}{8}\Delta^3 y_{x+\frac{1}{2}} + \ldots$$

Sol. L.H.S.
$$= \left(1 - \frac{1}{8}\frac{\Delta^2}{E} + \frac{1.3}{8.16}\frac{\Delta^4}{E^2} - \frac{1.3.5}{8.16.24}\frac{\Delta^6}{E^3} + \ldots\right) y_x$$
$$= \left(1 + \frac{1}{4}\frac{\Delta^2}{E}\right)^{-1/2} y_x = \left\{\frac{4E + (E-1)^2}{4E}\right\}^{-1/2} y_x$$
$$= 2E^{1/2}(1+E)^{-1} y_x$$
$$= 2E^{1/2}\{2+\Delta\}^{-1} y_x$$
$$= E^{1/2}\left(1 + \frac{\Delta}{2}\right)^{-1} y_x$$
$$= E^{1/2}\left\{1 - \frac{\Delta}{2} + \frac{1}{4}\Delta^2 - \frac{1}{8}\Delta^3 \ldots\right\} y_x$$
$$= y_{x+\frac{1}{2}} - \frac{1}{2}\Delta y_{x+\frac{1}{2}} + \frac{1}{4}\Delta^2 y_{x+\frac{1}{2}} \ldots\ldots$$
$$= \text{R.H.S.}$$

Ex. 4-37. *Show that* $y_0 + {}^x c_1 \Delta y_1 + {}^x c_2 \Delta^2 y_2 + {}^x c_3 \Delta^3 y_3 + \ldots$
$$= y_x + {}^x c_1 \Delta^2 y_{x-1} + {}^x c_2 \Delta^4 y_{x-2} + \ldots$$

Sol. R.H.S.
$$= \left(1 + {}^x c_1 \frac{\Delta^2}{E} + {}^x c_2 \frac{\Delta^4}{E^2} + \ldots\right) y_x$$
$$= \left(1 + \frac{\Delta^2}{E}\right)^x y_x$$
$$= E^{-x}\{E + \Delta(E-1)\}^x y_x$$
$$= E^{-x}\{E - \Delta + \Delta E\}^x y_x$$
$$= E^{-x}\{1 + \Delta E\}^x y_x$$
$$= E^{-x}\{1 + {}^x c_1 \Delta E + {}^x c_2 \Delta^2 E^2 + {}^x c_3 \Delta^3 E^3 + \ldots\} y_x$$
$$= y_0 + {}^x c_1 \Delta y_1 + {}^x c_2 \Delta^2 y_2 + {}^x c_3 \Delta^3 y_3 + \ldots$$
$$= \text{L.H.S.}$$

Ex. 4-38. *Show that*
$$\Delta^n y_{x-n} = y_x - {}^n c_1 y_{x-1} + {}^n c_2 y_{x-2} + \ldots + (-1)^n y_{x-n}$$

Sol. R.H.S.
$$= (1 - {}^n c_1 E^{-1} + {}^n c_2 E^{-2} \ldots) y_x$$
$$= (1 - E^{-1})^n y_x = E^{-n}(E-1)^n y_x$$
$$= E^{-n}\Delta^n y_x = \Delta^n y_{x-n}.$$

Ex. 4-39. *Show that*
$$U_0 + U_1 + \ldots + U_n = {}^{n+1}c_1 U_0 + {}^{n+1}c_2\ \Delta U_0 + \ldots \Delta^n U_0.$$

Sol. L.H.S.
$$= (1 + E + E^2 + \ldots + E^n) U_0$$
$$= \frac{E^{n+1} - 1}{E - 1} U_0$$
$$= \frac{1}{\Delta}\{(1+\Delta)^{n+1} - 1\} U_0$$
$$= \frac{1}{\Delta}\{{}^{n+1}c_1 \Delta + {}^{n+1}c_2 \Delta^2 + \ldots + \Delta^{n+1}\} U_0$$

$$= {}^{n+1}c_1\, U_0 + {}^{n+1}c_2\, \Delta\, U_0 + \ldots\ldots \Delta^n\, U_0.$$

Ex. 4-40. *Prove that*

$$\sum_{0}^{n-1} U_r x^r = \frac{U_0 - x^n U_n}{1-x} + \frac{x}{(1-x)^2}\,(\Delta U_0 - x^n\,\Delta U_n)$$

$$+ \frac{x^2}{(1-x)^3}\,(\Delta^2 U_0 - x^n\,\Delta^2 U_n) + \ldots.$$

Sol.

$$\begin{aligned}
\text{L.H.S.} &= U_0 + xU_1 + x^2U_2 + \ldots\ldots + x^{n-1}\,U_{n-1} \\
&= \{1 + xE + x^2E^2 + \ldots\, x^{n-1}\,E^{n-1}\}\, U_0 \\
&= \frac{1-x^nE^n}{1-xE}\, U_0 \\
&= (1 - x^n\,E^n)\,\frac{1}{1-x(1+\Delta)}\, U_0 \\
&= \frac{1-x^nE^n}{1-x}\left\{1 - \frac{x\Delta}{1-x}\right\}^{-1} U_0 \\
&= \frac{1-x^nE^n}{1-x}\left\{1 + \frac{x\Delta}{1-x} + \frac{x^2\Delta^2}{(1-x)^2} + \ldots\right\} U_0 \\
&= \frac{1-x^nE^n}{1-x}\left\{U_0 + \frac{x}{1-x}\,\Delta U_0 + \frac{x^2}{(1-x)^2}\,\Delta^2\, U_0 + \ldots\right\} \\
&= \frac{U_0 - x^nU_n}{1-x} + \frac{x}{(1-x)^2}\,\{\Delta U_0 - x^n\,\Delta U_n\} \\
&\quad + \frac{x^2}{(1-x)^3}\,\{\Delta^2\, U_0 - x^n\,\Delta^2\, U_n\} + \ldots. \\
&= \text{R.H.S.}
\end{aligned}$$

Ex. 4-41. *Show that*

$$\delta\{f(x)g(x)\} = \mu f(x)\,\delta g(x) + \mu g(x)\,\delta f(x).$$

Sol.

$$\begin{aligned}
\text{R.H.S.} &= \left\{\frac{E^{\frac{1}{2}} + E^{-\frac{1}{2}}}{2}\, f(x)\right\}\left\{\left(E^{\frac{1}{2}} - E^{-\frac{1}{2}}\right) g(x)\right\} \\
&\quad + \left\{\frac{E^{\frac{1}{2}} + E^{-\frac{1}{2}}}{2}\, g(x)\right\}\left\{\left(E^{\frac{1}{2}} - E^{-\frac{1}{2}}\right) f(x)\right\} \\
&= \frac{1}{2}\left\{f\left(x+\frac{h}{2}\right) + f\left(x-\frac{h}{2}\right)\right\}\left\{g\left(x+\frac{h}{2}\right) - g\left(x-\frac{h}{2}\right)\right\} \\
&\quad + \frac{1}{2}\left\{g\left(x+\frac{h}{2}\right) + g\left(x-\frac{h}{2}\right)\right\}\left\{f\left(x+\frac{h}{2}\right) - f\left(x-\frac{h}{2}\right)\right\} \\
&= f\left(x+\frac{h}{2}\right) g\left(x+\frac{h}{2}\right) - f\left(x-\frac{h}{2}\right) g\left(x-\frac{h}{2}\right)
\end{aligned}$$

=

Ex. 4-42. *Sum the series* 1

Sol. U_n = *n*th *term* = (2*n* –

Difference table is

x	l
1	
2	2
3	1:
4	3
5	7:

From Ex. 4-39.

$U_1 + U_2 + \ldots.\ U_n$

$\therefore$ S

4.3. Interpolation Formulae

(1) *When the values of the* ... *are used :*

4.3-1. Newton's divided diffe

$f(x)$

Derivation. Let $f(x)$ be a ... points $x_0, x_1 \ldots. x_n$ which are no

$f(x, x_0)$

$\therefore$ $f(x)$

$f(x, x_0, x_1)$

$\therefore$ $f(x, x_0)$

Similarly $f(x, x_0, x_1)$

$f(x, x_0\, x_1, \ldots\, x_{n-1})$

Multiplying eqs. by $(x -$. ... adding.

$f(x)$

where R

This formula, due to Ne ... formula. When the values of

$\ldots\ldots \Delta^n U_0.$

$\Delta U_0 - x^n \Delta U_n)$

$\Delta^2 U_n) + \ldots$

$x^{n-1} U_{n-1}$
$E^{n-1}) U_0$

U_0

U_0

$^2\Delta^2$ $-x)^2 + \ldots \} U_0$

$U_0 + \frac{x^2}{(1-x)^2} \Delta^2 U_0 + \ldots \}$

$\Delta U_0 - x^n \Delta U_n\}$

$U_n\} + \ldots$

$- E^{-\frac{1}{2}}\Big) g(x)\Big\}$

$\frac{1}{2} - E^{-\frac{1}{2}}\Big) f(x)\Big\}$

$\Big\}\Big\{g\left(x+\frac{h}{2}\right) - g\left(x-\frac{h}{2}\right)\Big\}$

$\Big)\Big\}\Big\{f\left(x+\frac{h}{2}\right) - f\left(x-\frac{h}{2}\right)\Big\}$

$\frac{h}{2}\Big) g\left(x-\frac{h}{2}\right)$

$$= \delta\{f(x)\,g(x)\} = \text{L.H.S.}$$

Ex. 4-42. *Sum the series* $1^3 + 3^3 + 5^3 + \ldots$ *upto n terms.*

Sol. $U_n = n$th *term* $= (2n-1)^3$

Difference table is

x	U_x	Δ	Δ^2	Δ^3
1	1			
		26		
2	27		72	
		98		48
3	125		120	
		218		48
4	343		168	
		386		
5	729			

From Ex. 4-39.

$$U_1 + U_2 + \ldots\, U_n = {}^nc_1\, U_1 + {}^nc_1\, \Delta U_1 + \ldots\ldots + \Delta^{n-1} U_1$$

$$\therefore \quad S = n + \frac{n(n-1)}{2!}\,26 + \frac{n(n-1)(n-2)}{3!}\,72 + \frac{n(n-1)(n-2)(n-3)}{4!}\,48$$

$$= n\{1 + 13\,(n-1) + 12(n^2 - 3n + 2) + 2(n^3 - 6n^2 + 11n - 6)\}$$

$$= n\{2n^3 - n\} = n^2\,(2n^2 - 1).$$

4.3. Interpolation Formulae

(1) *When the values of the argument are not equidistant, the following two formulae are used :*

4.3-1. Newton's divided difference Formula

$$f(x) = f(x_0) + (x-x_0)\,f(x_0, x_1) + (x-x_0)\,(x-x_1)\,f(x_0, x_1, x_2) + \ldots + (x-x_0)\,(x-x_1)\,\ldots\,(x-x_{n-1})\,f(x_0, x_1\,\ldots\, x_n)$$

Derivation. Let $f(x)$ be a function which takes n values $f(x_0), f(x_1)\,\ldots\, f(x_n)$ at the points $x_0, x_1 \ldots\, x_n$ which are not necessarily equidistant. Then by def

$$f(x, x_0) = \frac{f(x) - f(x_0)}{x - x_0}$$

$$\therefore \quad f(x) = f(x_0) + (x - x_0)\,f(x, x_0)$$

$$f(x, x_0, x_1) = \frac{f(x, x_0) - f(x_0, x_1)}{x - x_1}$$

$$\therefore \quad f(x, x_0) = f(x_0, x_1) + (x - x_1)\,f(x, x_0, x_1)$$

Similarly

$$f(x, x_0, x_1) = f(x_0, x_1, x_2) + (x - x_2)\,f(x, x_0, x_1, x_2)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$f(x, x_0\, x_1, \ldots\, x_{n-1}) = f(x_0, x_1, \ldots\, x_n) + (x - x_n)\,f(x, x_0, \ldots\, x_n)$$

Multiplying eqs. by $(x - x_0), (x - x_0)\,(x - x_1), \ldots\, (x - x_0)\,(x - x_1)\,\ldots\ldots\ldots\, (x - x_{n-1})$ and adding.

$$f(x) = f(x_0) + (x - x_0)\,f(x_0, x_1) + (x - x_0)\,(x - x_1)\,f(x_0, x_1, x_2) + \ldots + (x - x_0)\,(x - x_1)\,\ldots\,(x - x_{n-1})\,f(x_0, \ldots\, x_n) + R$$

where $\quad R = (x - x_0)\,(x - x_1)\,\ldots\,(x - x_n)\,f(x, x_0, \ldots\, x_n)$

This formula, due to Newton, is called Newton's divided difference interpolation formula. When the values of $f(x)$ for $x = x_0, x_1 \ldots\, x_n$ are known the evaluation of $f(x)$ is

reduced to the problem of evaluation R. If it is known or negligible, the required value of $f(x)$ can be calculated from above formula. In the case of a polynomial of nth degree, since $(n+1)$th order divided difference is zero, $R = 0$.

$\therefore$ If $f(x)$ is polynomial of nth degree.

$$f(x) = f(x_0) + (x - x_0) f(x_0, x_1) + (x - x_0)(x - x_1) f(x_0, x_1, x_2) + \dots + (x - x_0)(x - x_1) \dots (x - x_{n-1}) f(x_0, \dots x_n)$$

4.3-2. Lagrange's Formula

$$f(x) = \sum_{i=0}^{n} f(x_i) \left\{ \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)} \right\}$$

Drivation. Let $f(x)$ be a function which takes value $f(x_0), f(x_1), \dots f(x_n)$ for $(n+1)$ distinct points $x_0, x_1, \dots x_n$ and it is required to find a polynomial

$$P(x) = a_0 + a_1 x + \dots + a_n x^n$$

with the property that

$$P(x_i) = f(x_i) \qquad i = 0, 1, \dots n$$

The resulting polynomial is called Lagrange's interpolation polynomial or formula.

Evidently the unique polynomial $P(x)$ (of degree $\leq n$) with the required property is

$$P(x) = L_0(x) f(x_0) + L_1(x) f(x_1) + \dots + L_n(x) f_{x_n})$$

where $L_i(x)$ is a polynomial of degree n in x with the property that

$$L_i(x_j) = 0 \qquad j \neq i$$
$$= 1 \qquad j = i$$

Evidently $L_i(x)$ has the form

$$L_i(x) \equiv A_1(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)$$

$$\therefore \quad A_i = \frac{L_i(x_i)}{(x_i - x_0)(x_i - x_1)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)}$$

$$= \frac{1}{\prod\limits_{\substack{j \neq 0 \\ j \neq i}}^{n} (x_i - x_j)}$$

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)}$$

$$\therefore \quad P(x) = \sum_{i=0}^{n} f(x_i) \left\{ \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)} \right\}$$

Note. The polynomial $P(x)$ is unique because if $P(x)$ and $Q(x)$ be two such polynomials, then $\{P(x) - Q(x)\}$, a polynomial of degree $\leq n$, will have $(n+1)$ zeros $x_0, x_1, \dots x_n$ which is possible only when $P(x) \equiv Q(x)$.

Ex. 4-43. *Use Newton's formula for unequal intervals to find $f(8)$ from the following set of values :*

x :	4	5	7	10	11	13
$f(x)$:	2	4	8	104	114	452

Sol.

x	$f(x)$	1st order
4	2	
		2
5	4	
		2
7	8	
		32
10	104	
		10
11	114	
		169
13	452	

$\therefore \quad f(8) = 2 + ($

$+ (8 -$

$+ (8 \cdot$

Ex. 4-44. *Given the valu*

x : 4 5

$f(x)$: 48 100

form the table of divided diffe

Sol.

x	$f(x)$	1st ord
4	48	52
5	100	97
7	294	202
10	900	310
11	1210	409
13	2028	

$\therefore \quad f(2) = 48 + (2 - 4$

$f(15) = 48 + (15 -$

$+ (15 - 4)$

Ex. 4-45. *The function 3*

1, 2 and 4 respectively, Obtai

value differs from $3^3 = 27$.

MATHEMATICAL STATISTICS

ɔle, the required value of $f(x)$
ıial of nth degree, since $(n +$

$:-x_0)\ (x-x_1)\ f(x_0, x_1, x_2)$
$\text{c}-x_{n-1})\ f(x_0, \ldots. x_n)$

$), f(x_1), \ldots f(x_n)$ for $(n+1)$
al

$i = 0,\ 1, \ldots. n$
ı polynomial or formula.
the required property is
$+ L_n(x) f_{x_n})$
hat

$\text{ı}) (x - x_{i+1}) \ldots. (x - x_n)$
$)$
$\text{ı})(x_i - x_{i+1}) \ldots (x_i - x_n)$

be two such polynomials,
:eros $x_0, x_1, \ldots. x_n$ which is

d f(8) from the following

11	13
114	452

Sol.

Divided Difference Table

x	$f(x)$	*1st order*	*2nd order*	*3rd order*	*4th order*	*5th order*
4	2					
		2				
5	4		0	1		
		2			$-\frac{5}{12}$	
7	8		6	$-\frac{23}{12}$		$\frac{5}{24}$
		32			$\frac{35}{24}$	
10	104		$-\frac{11}{2}$			
		10		$\frac{117}{12}$		
11	114		53			
		169				
13	452					

$$\therefore \quad f(8) = 2 + (8-4)(2) + (8-4)(8-5)(0)$$
$$+ (8-4)(8-5)(8-7)(1) + (8-4)(8-5)(8-7)\ (8-10)\left(\frac{-5}{12}\right)$$
$$+ (8-4)(8-5)(8-7)(8-10)(8-11)\left(\frac{5}{24}\right) = 47.$$

Ex. 4-44. *Given the values*

x :	4	5	7	10	11	13
$f(x)$:	48	100	294	900	1210	2028

form the table of divided differences and use it to obtain $f(2)$ end $f(15)$.

Sol.

Divided Difference Table

x	$f(x)$	*1st order*	*2nd order*	*3rd order*	*4th order*	*5th order*
4	48					
		52				
5	100		15			
		97		1		
7	294		21		0	
		202		1		0
10	900		27		0	
		310		1		
11	1210		33			
		409				
13	2028					

$$\therefore \quad f(2) = 48 + (2-4)(52) + (2-4)(2-5)(15) + (2-4)(2-5)(2-7)(1) = 4$$
$$f(15) = 48 + (15-4)(52) + (15-4)(15-5)(15)$$
$$+ (15-4)(15-5)(15-7)(1) = 3150.$$

Ex. 4-45. *The function* 3^x *tables, as it should, the values* 1, 3, 9 *and* 81 *when x equals* 0, 1, 2 *and* 4 *respectively, Obtain the value corresponding to* $x = 3$ *and explain why the resulting value differs from* $3^3 = 27$.

Sol.

Divided Difference Table

x	3^x	1st order	2nd order	3rd order
0	1			
		2		
1	3		2	
		6		2
2	9		10	
		36		
4	81			

$$f(3) = 1 + (3-0)\,2 + (3-0)(3-1)(2) + (3-0)(3-1)(3-2)(2) = 31$$

Interpolating value differs from actual value because $f(x) = 3^x$ is not a polynomial.

Ex. 4-46. *The mode of a certain frequency curve $y = f(x)$ is very near $x = 9$ and the values of the frequency density for $x = 8{\cdot}9$, $9{\cdot}0$ and $9{\cdot}3$ are respectively equal to $0{\cdot}30$, $0{\cdot}35$ and $0{\cdot}25$. Calculate the approximate value of the mode.*

Sol.

Divided Difference Table

x	$f(x)$	1st order	2nd order
8·9	0·30		
		0·5	
9·0	0·35		– 2.08
		– 0·33	
9·3	0·25		

$$\therefore \quad f(x) = 0{\cdot}30 + (x - 8{\cdot}9)(0{\cdot}5) + (x - 8{\cdot}9)(x - 9{\cdot}0)(-2{\cdot}08)$$

$$\therefore \quad f'(x) = 0{\cdot}5 - (2{\cdot}08)\,\{(x - 9{\cdot}0) + (x - 8{\cdot}9)\}$$

For modal value of x, $f'(x) = 0$.

$$\therefore \quad 0{\cdot}5 - (2{\cdot}08)\,\{2x - 17{\cdot}9\} = 0$$

$$\therefore \quad 4{\cdot}16\,x = 0{\cdot}5 + 37{\cdot}232 = 37{\cdot}732$$

$$\therefore \quad x = 9{\cdot}07.$$

Ex. 4-47. *Use Lagrange's formula for interpolation to derive the form of the function $y = f(x)$, given*

x :	0	2	3	6
$f(x)$:	659	705	729	804

Sol.

$$f(x) = \frac{(x-2)(x-3)(x-6)}{(0-2)(0-3)(0-6)}(659) + \frac{(x-0)(x-3)(x-6)}{(2-0)(2-3)(2-6)}(705)$$

$$+ \frac{(x-0)(x-2)(x-6)}{(3-0)(3-2)(3-6)}(729) + \frac{(x-0)(x-2)(x-3)}{(6-0)(6-2)(6-3)}(804)$$

$$= -\frac{1}{72}x^3 + \frac{29}{72}x^2 + \frac{89}{4}x + 659.$$

Ex. 4-48. *Use Lagrange's formula to find $f(5)$ from the following data :*

x :	2	3	4	6	7
$f(x)$:	1	5	13	61	125

Sol.

$$f(5) = \frac{(5-3)(5-4)(5-6)(5-7)}{(2-3)(2-4)(2-6)(2-7)}(1) + \frac{(5-2)(5-4)(5-6)(5-7)}{(3-2)(3-4)(3-6)(3-7)}(5)$$

$$+ \frac{(5-2)(5-3)(5-6)(5-7)}{(4-2)(4-3)(4-6)(4-7)}(13) + \frac{(5-2)(5-3)(5-4)(5-7)}{(6-2)(6-3)(6-4)(6-7)}(61)$$

+ (5–2)(5– / (7–2)(7–

Ex. 4-49. *The following valu*
$f(1) =$

Find the value of $f(6)$ and al

Sol. $f(x) =$ (x–2)(x–7 / (1–2)(1–7

+ (x–1)(x– / (7–1)(7–

$= -\frac{1}{6}x^2 + \frac{3}{2}x$

$\therefore \quad f(6) = \frac{17}{3}$

$f'(x) = -\frac{1}{3}x + \frac{3}{2}$

Put $f'(x) = 0$

$\therefore \quad x = \frac{9}{2}.$

Since $f''(x) = -\frac{1}{3} < 0, f(x)$ i

Ex. 4-50. *The following tab*
months of life:

Age (in months)	0
Weight (in lbs)	7·5

Estimate the weight of the ba

Sol. $f(7) =$

=

Ex. 4-51. *The observed val*
four positions 3, 7, 9 and 10 of th
give for the value of the function

$$+\frac{(5-2)(5-3)(5-4)(5-6)}{(7-2)(7-3)(7-4)(7-6)}(125)=28\cdot 6.$$

Ex. 4-49. *The following values of the function $f(x)$ for values of x are given :*

$$f(1)=4,\ f(2)=5,\ f(7)=5,\ f(8)=4$$

Find the value of $f(6)$ and also the value of x for which $f(x)$ is maximum.

Sol.

$$\begin{aligned}
f(x) &= \frac{(x-2)(x-7)(x-8)}{(1-2)(1-7)(1-8)}\cdot 4+\frac{(x-1)(x-7)(x-8)}{(2-1)(2-7)(2-8)}\cdot 5\\
&\quad+\frac{(x-1)(x-2)(x-8)}{(7-1)(7-2)(7-8)}\cdot 5+\frac{(x-1)(x-2)(x-7)}{(8-1)(8-2)(8-7)}\cdot 4\\
&= -\frac{1}{6}x^2+\frac{3}{2}x+\frac{8}{3}.
\end{aligned}$$

$$\therefore \quad f(6) = \frac{17}{3}$$

$$f'(x) = -\frac{1}{3}x+\frac{3}{2}$$

Put $f'(x) = 0$

$$\therefore \quad x = \frac{9}{2}.$$

Since $f''(x)=-\frac{1}{3}<0, f(x)$ is max for $x=\frac{9}{2}$.

Ex. 4-50. *The following table gives the normal weights of babies during the first 12 months of life:*

Age (in months)	0	2	5	8	10	12
Weight (in lbs)	7·5	10·25	15	16	18	21

Estimate the weight of the baby at the age of 7 months.

Sol.

$$\begin{aligned}
f(7) &= \frac{(7-2)(7-5)(7-8)(7-10)(7-12)}{(0-2)(0-5)(0-8)(0-10)(0-12)}(7\cdot 5)\\
&\quad+\frac{(7-0)(7-5)(7-8)(7-10)(7-12)}{(2-0)(2-5)(2-8)(2-10)(2-12)}(10\cdot 25)\\
&\quad+\frac{(7-0)(7-2)(7-8)(7-10)(7-12)}{(5-0)(5-2)(5-8)(5-10)(5-12)}(15)\\
&\quad+\frac{(7-0)(7-2)(7-5)(7-10)(7-12)}{(8-0)(8-2)(8-5)(8-10)(8-12)}(16)\\
&\quad+\frac{(7-0)(7-2)(7-5)(7-8)(7-12)}{(10-0)(10-2)(10-5)(10-8)(10-12)}(18)\\
&\quad+\frac{(7-0)(7-2)(7-5)(7-8)(7-10)}{(12-0)(12-2)(12-5)(12-8)(12-10)}(21)\\
&= 15\cdot 67.
\end{aligned}$$

Ex. 4-51. *The observed values of a function are respectively* 168, 120, 72 *and* 63 *at four positions* 3, 7, 9 *and* 10 *of the independent variable. What is the best estimate you can give for the value of the function at the position* 6 *of the independent variable ?*

Sol.
$$f(6) = \frac{(6-7)(6-9)(6-10)}{(3-7)(3-9)(3-10)}(168) + \frac{(6-3)(6-9)(6-10)}{(7-3)(7-9)(7-10)}(120) + \frac{(6-3)(6-7)(6-10)}{(9-3)(9-7)(9-10)}(72)$$
$$+ \frac{(6-3)(6-7)(6-9)}{(10-3)(10-7)(10-9)}(63)$$
$$= 147.$$

Ex. 4-52. *Given the following table, find $\log_{10} 656$.*

x	:	654	658	659	661
$f(x) = \log_{10} x$	:	2·8156	2·8182	2·8189	2·8202

Sol.
$$f(656) = \frac{(656-658)(656-659)(656-661)}{(654-658)(654-659)(654-661)}(2{\cdot}8156)$$
$$+ \frac{(656-654)(656-659)(656-661)}{(658-654)(658-659)(658-661)}(2{\cdot}8182)$$
$$+ \frac{(656-654)(656-658)(656-661)}{(659-654)(659-658)(659-661)}(2{\cdot}8189)$$
$$+ \frac{(656-654)(656-658)(656-659)}{(661-654)(661-658)(661-659)}(2{\cdot}8202)$$
$$= 2{\cdot}81681 = 2{\cdot}8168.$$

Ex. 4-53. *Four equidistant values U_{-1}, U_0, U_1, and U_2 being given, a value is interpolated by Lagrange's formula. Show that it may be written in the form.*

$$U_x = yU_0 + xU_1 + \frac{y(y^2-1)}{3!}\Delta^2 U_{-1} + \frac{x(x^2-1)}{3!}\Delta^2 U_0$$

where $x + y = 1.$

Sol.
$$\text{R.H.S.} = (1-x)U_0 + xU_1 + \frac{(1-x)\{(1-x)^2-1\}}{3!}(E-1)^2 U_{-1}.$$
$$+ \frac{x(x^2-1)}{3!}(E-1)^2 U_0.$$
$$= (1-x)U_0 + xU_1 + \frac{(1-x)(x^2-2x)}{3!}(U_1 - 2U_0 + U_{-1})$$
$$+ \frac{x(x^2-1)}{3!}(U_2 - 2U_1 + U_0)$$
$$= \frac{x(1-x)(x-2)}{3!}U_{-1} + U_0\left\{(1-x) - \frac{1}{3}x(1-x)(x-2) + \frac{x(x^2-1)}{6}\right\}$$
$$+ U_1\left\{x + \frac{x(1-x)(x-2)}{6} - \frac{1}{3}x(x^2-1)\right\} + U_2\left\{\frac{x(x^2-1)}{3!}\right\}$$
$$= \frac{x(x-1)(x-2)}{(-1-0)(-1-1)(-1-2)}U_{-1} + \frac{(x+1)(x-1)(x-2)}{(0+1)(0-1)(0-2)}U_0$$

$+ \frac{(x+1)(}{(1+1)(}$

Ex. 4-54. *Given log* 100 = 2 2·0170, *find fog* 102.

Sol. log 102 =

=

(2) *When the values of the arg*

4.3-3 Newton's Forward Interp

$f(x) =$

It is used to interpolate near tl

Derivation. Let $y=f(x)$ be a fi values $x_0, x_1 \ldots\ldots x_n$ of x. Let

$I(x) =$

where the co-efficients $a_0, a_1, \ldots.$ a for $x = x_0, x_1, \ldots. x_n$ respectively. S

$x_1 - x_0 =$

$\therefore$ $x_i - x_0 =$

Now $I(x_0) =$

$I(x_1) =$

$\therefore$ $a_1 =$

$I(x_2) =$

$\therefore$ $a_2 =$

Similarly, $a_3 =$

$\therefore$ $I(x) =$

Let $U =$

Then

$$+\frac{(x+1)(x-0)(x-2)}{(1+1)(1-0)(1-2)}\,U_1+\frac{(x+1)(x-0)(x-1)}{(2+1)(2-0)(2-1)}\,U_2=U_x.$$

Ex. 4-54. *Given log* 100 = 2, *log* 101 = 2·0043, *log* 103 = 2·0128, *log* 104 = 2·0170, *find fog* 102.

Sol.

$$\begin{aligned}\log 102 &= \frac{(102-101)(102-103)(102-104)}{(100-101)(100-103)(100-104)}\,(2)\\ &\quad+\frac{(102-100)(102-103)(102-104)}{(101-100)(101-103)(101-104)}\,(2{\cdot}0043)\\ &\quad+\frac{(102-100)(102-101)(102-104)}{(103-100)(103-101)(103-104)}\,(2{\cdot}0128)\\ &\quad+\frac{(102-100)(102-101)(102-103)}{(104-100)(104-101)(104-103)}\,(2{\cdot}0170)\\ &= 2{\cdot}0086.\end{aligned}$$

(2) *When the values of the argument are equidistant the following formulae are used* :

4.3-3 Newton's Forward Interpolation Formula

$$f(x)=y_0+U^{(1)}\,\Delta y_0+\frac{U^{(2)}}{2!}\,\Delta^2 y_0+\ldots.+U^{(n)}\,\frac{\Delta^n y_0}{n!}$$

It is used to interpolate near the beginning of the table.

Derivation. Let $y=f(x)$ be a function which assumes the values $y_0, y_1 \ldots y_n$ for equidistant values $x_0, x_1 \ldots.. x_n$ of x. Let

$$\begin{aligned}I(x) &= a_0+a_1(x-x_0)+a_2(x-x_0)(x-x_1)\\ &\quad+\ldots.+\ldots..+a_n(x-x_0)(x-x_1)\ldots.(x-x_{n-1})\end{aligned}$$

where the co-efficients $a_0, a_1, \ldots. a_n$ are to be determined s.t. $I(x)$ takes the values $y_0, y_1 \ldots. y_n$ for $x=x_0, x_1, \ldots. x_n$ respectively. Since the values x_i are equidistant,

$$x_1-x_0 = x_2-x_1=\ldots\ldots=x_n-x_{n-1}=h \text{ (say)}$$

$$\therefore \quad x_i-x_0 = ih \quad i=1,\ldots.n$$

Now $\quad I(x_0) = a_0=y_0$

$$I(x_1) = a_0+a_1(x_1-x_0)=y_1$$

$$\therefore \quad a_1=\frac{y_1-y_0}{h}=\frac{1}{h}\,\Delta y_0$$

$$I(x_2)=a_0+a_1(x_2-x_0)+a_2(x_2-x_0)(x_2-x_1)=y_2$$

$$\therefore \quad a_2=\frac{1}{2h^2}\{y_2-2y_1+y_0\}=\frac{1}{2!h^2}\,\Delta^2 y_0$$

Similarly, $\quad a_3=\dfrac{\Delta^3 y_0}{3!h^3},\ldots..,a_n=\dfrac{1}{n!h^n}\,\Delta^n y_0$

$$\therefore \quad \begin{aligned}I(x) &= y_0+\frac{(x-x_0)}{h}\,\Delta y_0+\frac{(x-x_0)(x-x_1)}{2!h^2}\,\Delta^2 y_0\\ &\quad+\ldots.+\frac{(x-x_0)(x-x_1)\ldots.(x-x_{n-1})}{n!h^n}\,\Delta^n y_0\end{aligned}$$

Let $\quad U=\dfrac{x-x_0}{h}$

Then $\quad \dfrac{(x-x_0)(x-x_1)\ldots(x-x_{n-1})}{h^n}$

$$= \left(\frac{x-x_0}{h}\right)\left(\frac{x-x_0}{h}-1\right).....\left(\frac{x-x_0}{h}-\overline{n-1}\right)$$

$$= U(U-1)\;....\;(U-\overline{n-1}) = U^{(n)}$$

$$\therefore \qquad I(x) = y_0 + U^{(1)}\,\Delta y_0 + \frac{U^{(2)}}{2!}\,\Delta^2 y_0 + ... + \frac{U^{(n)}}{n!}\,\Delta^n y_0.$$

Ex. 4-55. *Given the following pairs of corresponding values of x and y.*

x :	20	25	30	35
y :	73	198	573	1198

Find the estimated value of y for x = 22.

Sol. **Difference Table**

x	y	Δy	$\Delta^2 y$	
20	73			
		125		
25	198		250	$U = \frac{x-20}{5}$
		375		
30	573		250	$= \frac{22-20}{5} = 0{\cdot}4$
		625		
35	1198			

$$\therefore \qquad U_{22} = 73 + (0{\cdot}4)\,(125) + \frac{(0{\cdot}4)(-0{\cdot}6)}{2!}\,(250)$$

$$= 93.$$

Ex. 4.56. *Given sin 45° = 0·7071, sin 50° = 0·7660, sin 55° = 0·8192 sin 60° = 0·8660. Find sin 48°.*

Sol.

Difference Table

x	$y = \sin x$	Δy	$\Delta^2 y$	$\Delta^3 y$	
45	0·7071				
		0·0589			
50	0·7660		– 0·0057		
		0·0532		– 0·0007	$U = \frac{48-45}{5} = 0{\cdot}6$
55	0·8192		– 0·0064		
		0·0468			
60	0·8660				

$$\therefore \qquad \sin 48^\circ = 0{\cdot}7071 + (0{\cdot}6)\,(0{\cdot}0589) + \frac{(0{\cdot}6)(-0{\cdot}4)}{2!}(-\,0{\cdot}0057)$$

$$+ \frac{(0{\cdot}6)(-0{\cdot}4)(-1{\cdot}4)}{3!}\,(-\,0{\cdot}0007)$$

$$= 0{\cdot}7431.$$

Ex. 4-57. *Find the number of men getting wages between Rs. 10 and Rs. 15 from the following table :*

Wages per week (in Rs.)	0 – 10	10 – 20	20 – 30	30 – 40
Frequency	9	30	35	42

Sol. Rewriting data in cumu

No. of persons getting less than	*Freq.*
10	9
20	39
30	74
40	116

∴ No. of men getting less tl

9 + (0·5) (30) +

+

∴ No. of men getting betw

=

Ex. 4-58. *Use Newton's forn from the table given below :*

Age	*Annual*
20	0
24	0

Sol.

Age	*Premiu*
20	0·0142
24	0·0158
28	0·017
32	0·019

∴ Premium at age 25 = 0·

$+ \frac{(1{\cdot}25)(0{\cdot}25)}{2!}$ (0·

= 0·01625.

Ex. 4-59. *The following examination.*

Not more than....
40
45
50
55

Sol. Rewriting data in cumulative frequency form and taking differences :

Difference Table

No. of persons getting less than	*Freq.*	Δ	Δ^2	Δ^3	
10	9				
		30			
20	39		5		
		35		2	$U = \frac{15-10}{10} = 0.5$
30	74		7		
		42			
40	116				

$\therefore$ No. of men getting less than Rs. 15 are

$$9 + (0.5)(30) + \frac{(0.5)(-0.5)}{2!}(5) + \frac{(0.5)(-0.5)(-1.5)}{3!}(2) = 23.5 \cong 24.$$

$\therefore$ No. of men getting between Rs. 10 and Rs. 15.

$$= 24 - 9 = 15.$$

Ex. 4-58. *Use Newton's formula for interpolation to find annual net premium at age 25 from the table given below :*

Age	*Annual Net premium*	*Age*	*Annual Net Premium*
20	0.01427	28	0.01772
24	0.01581	32	0.01996

Sol.

Difference Table

Age	*Premium*	Δ	Δ^2	Δ^3	
20	0.01427				
		0.00154			
24	0.01581		0.00037		
		0.00191		-0.00004	$U = \frac{25-20}{4} = 1.25$
28	0.01772		0.00033		
		0.00224			
32	0.01996				

$\therefore$ Premium at age 25 $= 0.01427 + (1.25)(0.00154)$

$$+ \frac{(1.25)(0.25)}{2!}(0.00037) + \frac{(1.25)(0.25)(-0.75)}{3!}(-0.00004)$$

$$= 0.01625.$$

Ex. 4-59. *The following are the marks obtained by 492 candidates in a certain examination.*

Not more than....	*Candidates*	*Not more than...*	*Candidates*
40	212	60	460
45	296	65	481
50	368	70	490
55	429	75	492

Find out the number of candidates who secured more than 42 but not more than 45 marks.

Sol.

Difference Table

x	*Freq.*	Δ	D^2	Δ^3	Δ^4	Δ^5	Δ^6	Δ^7
40	212							
		84						
45	296		−12					
		72		1				
50	368		−11		−20			
		61		−19		59		
55	429		−30		39		−120	
		31		20		−61		210
60	460		−10		−22		90	
		21		−2		29		
65	481		−12		7			
		9		5				
70	490		−7					
		2						
75	492							

$$U = \frac{42-40}{5} = 0{\cdot}4$$

$\therefore$ Number of candidates getting marks less than 42

$$= 212 + (0{\cdot}4)(84) + \frac{(0{\cdot}4)(-0{\cdot}6)}{2!}(-12) + \frac{(0{\cdot}4)(-0{\cdot}6)(-1{\cdot}6)}{3!}(1)$$

$$+ \frac{(0{\cdot}4)(-0{\cdot}6)(-1{\cdot}6)(-2{\cdot}6)}{4!}(-20)$$

$$+ \frac{(0{\cdot}4)(-0{\cdot}6)(-1{\cdot}6)(-2{\cdot}6)(-3{\cdot}6)}{5!}(59)$$

$$+ \frac{(0{\cdot}4)(-0{\cdot}6)(-1{\cdot}6)(-2{\cdot}6)(-3{\cdot}6)(-4{\cdot}6)}{6!}(-120)$$

$$+ \frac{(0{\cdot}4)(-0{\cdot}6)(-1{\cdot}6)(-2{\cdot}6)(-3{\cdot}6)(-4{\cdot}6)(-5{\cdot}6)}{7!}(210)$$

$= 256$ (approx.)

$\therefore$ Number of candidates getting marks more than 42 but not more than 45 $= 296 - 256$
$= 40.$

Ex. 4-60. *Find* $f(0{\cdot}0477)$ *from the following data :*

x	:	0	0·05	0·10	0·15	0·20
$f(x)$	:	1·00000	0·99750	0·99005	0·97775	0·96079

Sol.

Difference Table

x	$f(x)$	Δ	Δ^2	Δ^3	Δ^4
0·00	1·00000				
		− 0·00250			
0·05	0·99750		− 0·00495		
		− 0·00745		0·00010	
0·10	0·99005		− 0·00485		0·00009
		− 0·01230		0·00019	
0·15	0·97775		− 0·00466		
		− 0·01696			
0·20	0·96079				

$U =$

$\therefore \quad f(0{\cdot}0477) =$

$=$

Ex. 4-61. *Given*

$\sum_{1}^{10} f(x) =$

$\sum_{7}^{10} f(x) =$

find $f(1)$.

Sol. Let $s_x =$

Then

x	S_x
1	50042
4	32924
7	17521
10	4036

U

for S_2, x

$\therefore \quad S_2$

$\therefore \quad S_1 - S_2$

$\Rightarrow \quad f(1)$

4.3-4. Newton's Backward I

$f(x) = y_n + U^{(1)} \nabla y_n$

It is used to interpolate ne

Derivation. Let $y = f(x)$ b

values $x_0, x_1, \ldots, x_n$ of x. Let

$$U = \frac{0\cdot0477-0}{0\cdot05} = 0\cdot954$$

$$\begin{aligned}\therefore \quad f(0\cdot0477) &= 1\cdot00000 + (0\cdot954)(-0\cdot00250) \\ &\quad + \frac{(0\cdot954)(-0\cdot046)}{2!}(-0\cdot00495) \\ &\quad + \frac{(0\cdot954)(-0\cdot046)(-1\cdot046)}{3!}(0\cdot00010) \\ &\quad + \frac{(0\cdot954)(-0\cdot046)(-1\cdot046)(-2\cdot046)}{4!}(0\cdot00009) \\ &= 0\cdot9977240 \simeq 0\cdot99772.\end{aligned}$$

Ex. 4-61. *Given*

$$\sum_{1}^{10} f(x) = 500424, \quad \sum_{4}^{10} f(x) = 329240$$

$$\sum_{7}^{10} f(x) = 175212 \text{ and } f(10) = 40365,$$

find $f(1)$.

Sol. Let $\qquad s_x = \sum_{x}^{10} f(x), \qquad 1 \le x \le 10$

Then

x	S_x	ΔSx	$\Delta^2 Sx$	$\Delta^3 Sx$
1	500424			
		−171184		
4	329240		17156	
		− 154028		2025
7	175212		19181	
		− 134847		
10	40365			

$$U = \frac{x-1}{3}$$

for S_2, $\qquad x = 2$ and hence $U = \frac{1}{3}$

$$\begin{aligned}\therefore \quad S_2 &= 500424 + \frac{1}{3}(-171184) \\ &\quad + \frac{\left(\frac{1}{3}\right)\left(-\frac{2}{3}\right)}{2!}(17156) + \frac{\left(\frac{1}{3}\right)\left(\frac{-2}{3}\right)\left(\frac{-5}{3}\right)}{3!}(2025) \\ &= S_1 - 58842\cdot55 \\ \therefore \quad S_1 - S_2 &= 58842\cdot55 \\ \Rightarrow \quad f(1) &= 58842\cdot55.\end{aligned}$$

4.3-4. Newton's Backward Interpolation Formula

$$f(x) = y_n + U^{(1)}\nabla y_n + \frac{U^{(2)}}{2!}\nabla^2 y_n + \ldots + \frac{U^{(n)}}{n!}\nabla^n y_n$$

It is used to interpolate near the end of the table.

Derivation. Let $y = f(x)$ be a function which assumes the values $y_0, y_1 \ldots y_n$ for equidistant values $x_0, x_1, \ldots, x_n$ of x. Let

$I(x) = a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1}) + \ldots + a_n(x - x_n)(x - x_{n-1}) \ldots (x - x_1)$ where the co-efficients $a_0, a_1, \ldots a_n$ are to be determined *s.t.* $I(x)$ takes the values $y_0, y_1 \ldots y_n$ for $x = x_0, x_1, \ldots x_n$ respectively.

$$\begin{aligned}
\text{Let} \quad x_1 - x_0 &= x_2 - x_1 = \ldots = x_n - x_{n-1} = h \\
\therefore \quad x_i - x_0 &= ih \qquad i = 1, \ldots n \\
\text{Now} \quad I(x_n) &= a_0 = y_n \\
I(x_{n-1}) &= a_0 + a_1(x_{n-1} - x_n) \\
\therefore \quad a_1 &= \frac{1}{h}(y_n - y_{n-1}) = \frac{1}{h}\nabla y_n \\
I(x_{n-2}) &= a_0 + a_1(x_{n-2} - x_n) + a_2(x_{n-2} - x_n)(x_{n-2} - x_{n-1}) \\
\therefore \quad a_2 &= \frac{1}{2h^2}\{y_n - 2y_{n-1} + y_{n-2}\} = \frac{1}{2!h^2}\nabla^2 y_n \\
\text{Similarly} \quad a_3 &= \frac{1}{3!h^3}\nabla^3 y_n, \ldots a_n = \frac{1}{n!h^n}\nabla^n y_n \\
\therefore \quad I(x) &= y_n + \frac{(x - x_n)}{h}\nabla y_n + \frac{(x - x_n)(x - x_{n-1})}{2!h^2}\nabla^2 y_n \\
&\quad + \ldots + \frac{(x - x_n)\ldots(x - x_1)}{n!h^n}\nabla^n y_n \\
\text{Let} \quad U &= \frac{x - x_n}{h}, \qquad \text{i.e., } x = x_n + Uh \\
\text{Then} \quad (x - x_n)\ldots(x - x_1) &= h^n U(U+1)\ldots(U + \overline{n-1}) \\
&= h^n U^{(n)} \\
\text{where} \quad U^{(n)} &= U(U+1)\ldots(U + n - 1) \\
\therefore \quad I(x) &= y_n + U^{(1)}\nabla y_n + \frac{U^{(2)}}{2!}\nabla^2 y_n + \ldots + \frac{U^{(n)}}{n!}\nabla^n y_n \\
\text{where} \quad U^{(n)} &= U(U+1)\ldots(U + \overline{n-1}).
\end{aligned}$$

Ex. 4-62. *Estimate the population in 1995 of a place having the following record.*

Year	1961	1971	1981	1991	2001
Population (in thousands)	46	66	81	93	101

Sol. Since 1975 is near the end of the table. Newton's backward formula will be used.

Difference Table

Year	*Population*	Δ	Δ^2	Δ^3	Δ^4	
1961	46					
		20				
1971	66		–5			
		15		2		
1981	81		– 3		–3	$U = \frac{1995 - 2001}{10} = -0.6$
		12		–1		
1991	93		–4			
		8				
2001	101					

$\therefore$ Population in 1995 $= 101 + (-0.6)(8) + \dfrac{(-0.6)(0.4)}{2!}(-4)$

$-x_n)(x-x_{n-1}) \ldots\ldots (x-x_1)$ where takes the values $y_0, y_1 \ldots\ldots y_n$ for

$= h$

$_{n-2}-x_n)(x_{n-2}-x_{n-1})$

$= \frac{1}{2!h^2}\nabla^2 y_n$

$\nabla^n y_n$

$\frac{x_n)(x-x_{n-1})}{2!h^2}\nabla^2 y_n$

$\nabla^n y_n$

$+ Uh$

$1)$

$_n+\ldots\ldots+\frac{U^{(n)}}{n!}\nabla^n y_n$

aving the following record.

1991	2001
93	101

ackward formula will be used.

Δ^4	
-3	$U=\frac{1995-2001}{10}$ $=-0{\cdot}6$

(-4)

$$+\frac{(-0{\cdot}6)(0{\cdot}4)(1{\cdot}4)}{3!}(-1)+\frac{(-0{\cdot}6)(0{\cdot}4)(1{\cdot}4)(2{\cdot}4)}{4!}(-3)$$

$$= 101-4{\cdot}8+0{\cdot}48+0{\cdot}056+0{\cdot}1008 = 96{\cdot}8368 = 96{\cdot}84 \text{ thousands.}$$

4.3-5. Central Difference Formulae

(a)
$$f(x) = y_0 + U\mu\delta y_0 + \frac{U^2}{2!}\delta^2 y_0 + \frac{U(U^2-1)}{3!}\mu\delta^3 y_0 + \frac{U^2(U^2-1)}{4!}\delta^4 y_0 + \ldots + \frac{U^2(U^2-1^2)\ldots(U^2-(r-1)^2\}}{2r!}\delta^{2r} y_0 + \frac{U(U^2-1^2)\ldots(U^2-r^2)}{(2r+1)!}\mu\delta^{2r+1} y_0 + \ldots \text{ (Stirling's formula)}$$

(b)
$$f(x) = \mu y_{\frac{1}{2}} + V\delta y_{\frac{1}{2}} + \frac{1}{2!}\left(V^2-\frac{1}{4}\right)\mu\delta^2 y_{\frac{1}{2}} + \frac{1}{3!}V\left(V^2-\frac{1}{4}\right)\delta^3 y_{\frac{1}{2}} + \ldots + \frac{1}{2r!}\left(V^2-\frac{1}{4}\right)\left(V^2-\frac{9}{4}\right)\ldots \left\{V^2-\frac{(2r-1)^2}{4}\right\}\mu\delta^{2r} y_{\frac{1}{2}} + \frac{1}{(2r+1)!}V\left(V^2-\frac{1}{4}\right)\left(V^2-\frac{9}{4}\right)\ldots \left\{V^2-\frac{(2r-1)^2}{4}\right\}\delta^{2r+1} y_{\frac{1}{2}} + \ldots\ldots\ldots \text{ (Bessel's formula)}$$

where
$$V = U - \frac{1}{2}$$

(c)
$$f(x) = y_0 + U^{(1)}\delta y_{-\frac{1}{2}} + \frac{(U+1)^{(2)}}{2!}\delta^2 y_0 + \ldots \frac{(U+r)^{(2r)}}{2r!}\delta^{2r} y_0 + \frac{(U+r)^{(2r+1)}}{(2r+1)!}\delta^{2r+1} y_{-\frac{1}{2}} + \ldots\ldots \text{ (Gauss backward formula)}$$

It is used when U is negative.

(d)
$$f(x) = y_0 + U^{(1)}\delta y_{\frac{1}{2}} + \frac{U^{(2)}}{2!}\delta^2 y_0 + \ldots + \frac{(U+\overline{r-1})^{(2r-1)}}{(2r-1)!}\delta^{2r-1} y_{\frac{1}{2}} + \frac{(U+\overline{r-1})^{(2r)}}{2r!}\delta^{2r} y_0 + \ldots \text{ (Gauss forward formula)}$$

It is used when U is positive.
Central difference formulae are used to interpolate near the middle of the table.

Ex. 4-63. *Find the value of an annuity at $5\frac{3}{8}$% from the following table :*

Rate per cent	*Annuity-value*	*Rate per cent*	*Annuity-value*
4	17·29203	$5\frac{1}{2}$	14·53375
$4\frac{1}{2}$	16·28889	6	13·76483
5	15·37245		

Sol. Since $5\frac{3}{8}\%$ lies near the middle of the table, any central difference formula can be applied.

Difference Table

Rate	*Annuity-value*	Δ	Δ^2	Δ^3	Δ^4
4·0	17·29203				
		− 1·00314			
4.5	16·28889		0·08670		
		− 0·91644		−0·00896	
5·0	15·37245		0·07774		0·00100
		− 0·83870		− 0·00796	
5·5	14·53375		0·06978		
		− 0·76892			
6·0	13·76483				

Here $x_0 = 5{\cdot}0; \; h = 0{\cdot}5$

$$\therefore \quad U = \frac{5\frac{3}{8} - 5}{0\cdot5} = 0\cdot75$$

(*i*) **Using Stirling's formula.**

$$y_x = y_0 + U\mu\delta y_0 + \frac{U^2}{2!}\delta^2 y_0 + \frac{U(U^2-1)}{3!}\mu\delta^3 y_0 + \frac{U^2(U^2-1)}{4!}\delta^4 y_0 + \ldots$$

$$= y_0 + U\mu\Delta y_{-\frac{1}{2}} + \frac{U^2}{2!}\Delta^2 y_{-1} + \frac{U(U^2-1)}{3!}\mu\Delta^3 y_{-3/2} + \frac{U^2(U^2-1)}{4!}\Delta^4 y_{-2}\ \ldots\ldots$$

$$= y_0 + U\left\{\frac{\Delta y_0 + \Delta y_{-1}}{2}\right\} + \frac{U^2}{2!}\Delta^2 y_{-1} + \frac{U(U^2-1)}{3!}\left\{\frac{\Delta^3 y_{-1} + \Delta^3 y_{-2}}{2}\right\} + \frac{U^2(U^2-1)}{4!}\Delta^4 y_{-2} + \ldots\ldots$$

$$= 15{\cdot}37245 + (0{\cdot}75)\left\{\frac{-0\cdot83870 - 0\cdot91644}{2}\right\} + \frac{(0\cdot75)^2}{2!}(0{\cdot}07774) + \frac{(0\cdot75)\{(0\cdot75)^2-1\}}{3!}\left\{\frac{-0\cdot00796 - 0\cdot00896}{2}\right\} + \frac{(0\cdot75)^2\{(0\cdot75)^2-1\}}{4!}(0{\cdot}00100)$$

$$= 14{\cdot}736589 \cong 14{\cdot}73659.$$

(*ii*) **Using Bessel's formula.**

$$y_x = \mu y_{1/2} + v\delta y_{1/2} + \frac{1}{2!}\left(V^2 - \frac{1}{4}\right)\mu\delta^2 y_{1/2} + \frac{1}{3!}V\left(V^2 - \frac{1}{4}\right)\delta^3 y_{1/2} + \frac{1}{4!}V\left(V^2 - \frac{1}{4}\right)\left(V^2 - \frac{9}{4}\right)\mu\delta^4 y_{1/2} + \ldots$$

where $V = U - \frac{1}{2}$

$= \frac{y_0 + y_1}{2}$

$+ \frac{1}{6}V($

$+ \frac{1}{24}($

Here $V = 0{\cdot}75 - ($

$\therefore \quad y_x = \frac{15\cdot372}{}$

$+ \frac{1}{2}\{($

$+ \frac{1}{2}(0\cdot$

$+ \frac{1}{24}\{$

$= 14{\cdot}953$

$= 14{\cdot}736.$

Ex. 4-64. *From the follow*

Age 2

Annual Premium (in Rs.) 28

Sol.

Age x	*u*	*Premium y*
24	−2	28·06
28	−1	30·19
32	0	32·75
36	1	35·94
40	2	40·00

By Stirling's formula, we

$y_{0\cdot25} = 32{\cdot}75$

$+ \frac{(0\cdot}{}$

$+ \frac{(0\cdot}{}$

where $$V = U - \frac{1}{2}$$

$$= \frac{y_0 + y_1}{2} + V\Delta y_0 + \frac{1}{2!}\left(V^2 - \frac{1}{4}\right)\left(\frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2}\right)$$

$$+ \frac{1}{6}V\left(V^2 - \frac{1}{4}\right)\Delta^3 y_{-1}$$

$$+ \frac{1}{24}\left(V^2 - \frac{1}{4}\right)\left(V^2 - \frac{9}{4}\right)\left(\frac{\Delta^4 y_{-1} + \Delta^4 y_{-2}}{2}\right) + \ldots$$

Here $$V = 0.75 - 0.5 = 0.25$$

$$\therefore \quad y_x = \frac{15.37245 + 14.53375}{2} + (0.25)(-0.83870)$$

$$+ \frac{1}{2}\left\{(0.25)^2 - 0.25\right\}\left\{\frac{0.06978 + 0.07774}{2}\right\}$$

$$+ \frac{1}{2}(0.25)\{(0.25)^2 - 0.25\}(-0.00796)$$

$$+ \frac{1}{24}\{(0.25)^2 - 0.25\}\{(0.25)^2 - 2.25\}(0.00100)$$

$$= 14.95310 - 0.209675 - 0.0069150 + 0.0000622 + 0.0000171$$

$$= 14.736589 = 14.73659.$$

Ex. 4-64. *From the following data, find the annual premium at the age of* 33.

Age	24	28	32	36	40
***Annual Premium* (*in Rs.*)**	28·06	30·19	32·75	35·94	40·00

Sol.

Age x	*u*	*Premium y*	Δ^1	Δ^2	Δ^3	Δ^4	
24	−2	28·06					
			2·13				
28	−1	30·19		0·43			
			2·56		0·20		
32	0	32·75		0·63		0·04	$U = \frac{33-32}{4} = 0.25$
			3·19		0·24		
36	1	35·94		0·87			
			4·06				
40	2	40·00					

By Stirling's formula, we have

$$y_{0.25} = 32.75 + (0.25)\left\{\frac{3.19 + 2.56}{2}\right\} + \frac{(0.25)^2}{2!}(0.63)$$

$$+ \frac{(0.25)\{(0.25)^2 - 1\}}{3!}\left\{\frac{0.24 + 0.20}{2}\right\}$$

$$+ \frac{(0.25)^2\{(0.25)^2 - 1\}}{4!}(0.04)$$

$$= 32{\cdot}75 + 0{\cdot}719 + 0{\cdot}02 - 0{\cdot}009 - 0{\cdot}000$$
$$= 33{\cdot}48.$$

Ex. 4-65. *Show that*

$$y_{2\frac{1}{2}} = \frac{1}{2}c + \frac{25(c-b)+3(a-c)}{256}$$

where $a = y_0 + y_5,\ b = y_1 + y_4, = c = y_2 + y_3$ if $\Delta^5 y_x$ are constant.

Sol. From Bessel's formula

$$y_x = \frac{\gamma_0+\gamma_1}{2} + V\Delta\gamma_0 + \frac{1}{2!}\left(V^2-\frac{1}{4}\right)\left(\frac{\Delta^2\gamma_0+\Delta^2\gamma_{-1}}{2}\right)$$
$$+ \frac{1}{6}V\left(V^2-\frac{1}{4}\right)\Delta^3\gamma_{-1} + \frac{1}{24}\left(V^2-\frac{1}{4}\right)\times$$
$$\left(V^2-\frac{9}{4}\right)\left(\frac{\Delta^4\gamma_{-1}+\Delta^4\gamma_{-2}}{2}\right)$$
$$+ \frac{1}{120}V\left(V^2-\frac{1}{4}\right)\left(V^2-\frac{9}{4}\right)\Delta^5\gamma_{-2}$$

Let $x_0 = 2$

Then $\gamma_0 = y_2,\ \gamma_{-1} = y_1,\ \gamma_{-2} = y_0,\ \gamma_1 = y_3,\ \gamma_2 = y_4$ and $\gamma_3 = y_5$.

Also $V = U - \frac{1}{2} = (x-x_0) - \frac{1}{2} = \left(2\frac{1}{2}-2\right) - \frac{1}{2} = 0$

$$\therefore\quad y_{2\frac{1}{2}} = \gamma_{\frac{1}{2}} = \frac{\gamma_0+\gamma_1}{2} - \frac{1}{16}\{(E-1)^2\gamma_0 + (E-1)^2\gamma_{-1}\}$$
$$+ \frac{9}{768}\{(E-1)^4\gamma_{-1} + (E-1)^4\gamma_{-2}\}$$
$$= \frac{\gamma_0+\gamma_1}{2} - \frac{1}{16}\{(\gamma_2 - 2\gamma_1 + \gamma_0) + (\gamma_1 - 2\gamma_0 + \gamma_{-1})\}$$
$$+ \frac{9}{768}\{(\gamma_3 - 4\gamma_2 + 6\gamma_1 - 4\gamma_0 + \gamma_{-1}) + (\gamma_2 - 4\gamma_1 + 6\gamma_0 - 4\gamma_{-1} + \gamma_{-2})\}$$
$$= \frac{y_2+y_3}{2} - \frac{1}{16}\{y_4 - y_3 - y_2 + y_1\}$$
$$+ \frac{9}{768}\{y_5 - 3y_4 + 2y_3 + 2y_2 - 3y_1 + y_0\}$$
$$= \frac{c}{2} - \frac{1}{16}(b-c) + \frac{3}{256}(a - 3b + 2c)$$
$$= \frac{c}{2} + \frac{1}{256}\{3(a-c) + 25(c-b)\}.$$

Ex. 4-66. *If third differences are constant, show that*

$$y_{x+\frac{1}{2}} = \frac{1}{2}(y_x + y_{x+1}) - \frac{1}{16}(\Delta^2 y_{x-1} + \Delta^2 y_x).$$

Sol. From Bessel's formula

$$\gamma_x = \frac{\gamma_0+\gamma_1}{2} + V\Delta\gamma_0 + \frac{1}{2!}\left(V^2-\frac{1}{4}\right)\left(\frac{\Delta^2\gamma_0+\Delta^2\gamma_{-1}}{2}\right)$$

$$+ \frac{1}{3!}V\Big($$

Let $x_0 = x$

Then $\gamma_0 = y_x,\ \gamma_1 = y_x$

$U = \left(x+\frac{1}{2}\right) -$

and $V = U - \frac{1}{2} =$

$\therefore\quad y_{x+\frac{1}{2}} = \frac{y_x + y_{x+}}{2}$

Ex. 4-67. *The following tabl of workers in a big manufacturin*

Earnings per month

up to Rs. 10

up to Rs. 20

up to Rs. 30

Find out the number of worl

Sol.

x	u	y	
10	– 2	50	y_{-2}
20	– 1	150	y_{-1}
30	0	300	y_0
40	1	500	y_1
50	2	700	y_2
60	3	800	y_3

$U(\text{for } x = 25) = \frac{25-30}{10}$

$U(\text{for } x = 35) = \frac{35-30}{10}$

For $x = 25$ since U is nega

$\therefore\quad y_{-0\cdot5} = 300 +$

$+ \frac{(-0}{}$

$= 300 -$

For $x = 35$, since U is pos

have

$y_{0\cdot5} = 300 +$

$+ \frac{(0}{}$

)0

$\Delta^5 y_x$ are constant.

$\dfrac{\gamma_0+\Delta^2\gamma_{-1}}{2}\Big)$

$\Big) \times$

and $\gamma_3 = y_5$.

$= 0$

$-1)^2\gamma_{-1}\}$

$\gamma_0+\gamma_{-1})\}$

$(\gamma_2 - 4\gamma_1 + 6\gamma_0 - 4\gamma_{-1} + \gamma_{-2})$

$)\}$

$\dfrac{+\Delta^2\gamma_{-1}}{2}\Big)$

$$+\frac{1}{3!}V\left(V^2-\frac{1}{4}\right)\Delta^3\gamma_{-1}$$

Let $\quad x_0 = x$

Then $\quad \gamma_0 = y_x,\ \gamma_1 = y_{x+1},\ \gamma_{-1} = y_{x-1}$, *etc.*

$$U = \left(x+\frac{1}{2}\right)-x=\frac{1}{2}$$

and $$V = U-\frac{1}{2}=0$$

$$\therefore \quad y_{x+\frac{1}{2}} = \frac{y_x+y_{x+1}}{2}-\frac{1}{16}(\Delta^2 y_x+\Delta^2 y_{x-1}).$$

Ex. 4-67. *The following table relates to income earned per month by a certain number of workers in a big manufacturing concern.*

Earnings per month	*Freq.*	*Earnings per month*	*Freq.*
up to Rs. 10	50	*up to Rs.* 40	500
up to Rs. 20	150	*up to Rs.* 50	700
up to Rs. 30	300	*up to Rs.* 60	800

Find out the number of workers falling within the Rs. 25-35 earning group.

Sol.

x	u	y		Δ^1	Δ^2	Δ^3	Δ^4	Δ^5	
10	-2	50	y_{-2}						
				100					
20	-1	150	y_{-1}		50				
				150		0			
30	0	300	y_0		50		-50		
				200		-50		0	$U=\dfrac{x-30}{10}$
40	1	500	y_1		0		-50		
				200		-100			
50	2	700	y_2		-100				
				100					
60	3	800	y_3						

$$U(\text{for } x=25)=\frac{25-30}{10}=-0{\cdot}5$$

$$U(\text{for } x=35)=\frac{35-30}{10}=0{\cdot}5.$$

For $x = 25$ since U is negative, we apply Gauss's backward formula. By this formula.

$$\therefore \quad y_{-0\cdot5} = 300+(-0{\cdot}5)(150)+\frac{(-0{\cdot}5)(0{\cdot}5)}{2!}(50)$$
$$+\frac{(-0{\cdot}5)\{(-0{\cdot}5)^2-1\}}{3!}(0)+\frac{(-0{\cdot}5)\{(-0{\cdot}5)^2-1)\}\{1{\cdot}5\}}{4!}(-50)$$
$$= 300-75-6{\cdot}25-1{\cdot}17=217{\cdot}58=218.$$

For $x = 35$, since U is positive, we apply Gauss's forward formula. By this formula we have

$$y_{0\cdot5} = 300+(0{\cdot}5)(200)+\frac{(0{\cdot}5)(-0{\cdot}5)}{2!}(50)$$
$$+\frac{(0{\cdot}5)\{(0{\cdot}25-1\}}{3!}(-50)+\frac{(0{\cdot}5)(0{\cdot}25-1)\{-1{\cdot}5\}}{4!}(-50)$$

$= 300 + 100 - 6{\cdot}25 + 3{\cdot}125 - 1{\cdot}17$
$= 395{\cdot}705 = 396.$

$\therefore$ No. of persons earning between Rs. 25 and Rs. 35 $= 396 - 218 = 178.$

Ex. 4-68. *If p, q, r, s, be the successive entries corresponding to equidistant arguments in a table, show that when third differences are taken into account, the entry corres-ponding to the argument half way between the arguments of q and r is $A + \frac{1}{24} B$, where A is the A.M. of q and r and B is the A.M. of $3q - 2p - s$ and $3r - 2s - p$.*

Sol. In Ex. 4-66, let

$$y_{x-1} = p,\ y_x = q,\ y_{x+1} = r \text{ and } y_{x+2} = s$$

Then
$$y_{x+\frac{1}{2}} = \frac{q+r}{2} - \frac{1}{16}\{(E-1)^2 y_x + (E-1)^2 y_{x-1}\}$$
$$= \frac{q+r}{2} - \frac{1}{16}\{y_{x+2} - y_{x+1} - y_x + y_{x-1}\}$$
$$= A - \frac{1}{16}\{s - r - q + p\}$$

Also
$$B = \frac{1}{2}\{(3q - 2p - s) + (3r - 2s - p)\}$$
$$= \frac{3}{2}(q + r - p - s)$$

$$\therefore \quad y_{x+\frac{1}{2}} = A + \frac{1}{24} B.$$

To find Missing Terms

Ex. 4-69. *Given* $U_0 + U_8 = 1{\cdot}9243$, $U_1 + U_7 = 1{\cdot}9590$, $U_2 + U_6 = 1{\cdot}9823$ *and* $U_3 + U_5 = 1{\cdot}9956$. *Find the* U_4.

Sol. Since eight entries are given, from these values a polynomial of degree seven can be obtained and hence $\Delta^7 U_x$ is assumed to be constant and consequently $\Delta^8 U_x = 0$ for all x.

$$\therefore \quad \Delta^8 U_0 = 0$$
i.e., $(E-1)^8 U_0 = 0$
i.e., $(E^8 - 8E^7 + 28E^6 - 56E^5 + 70E^4 - 56E^3 + 28E^2 - 8E + 1)\, U_0 = 0$
i.e., $(U_8 + U_0) - 8(U_7 + U_1) + 28\,(U_6 + U_2) - 56\,(U_5 + U_3) + 70U_4 = 0$

$$\therefore \quad 70U_4 = -1{\cdot}9243 + 8\,(1{\cdot}9590) - 28(1{\cdot}9823) + 56(1{\cdot}9956)$$
$$= 69{\cdot}9969$$
$$\therefore \quad U_4 = \frac{69{\cdot}9969}{70} = 0{\cdot}999956 \cong 1{\cdot}0000.$$

Ex. 4-70. *Find the missing terms in the following data :*

x :	2·0	2·1	2·2	2·3	2·4	2·5	2·6
y :	13·5	?	11·1	10	?	8·2	7·4

Sol. Taking the missing entries as x and y the difference table is given below :

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5
2·0	13·5					
		$x - 13{\cdot}5$				
2·1	x		$24{\cdot}6 - 2x$			
		$11{\cdot}1 - x$		$3x - 36{\cdot}8$		
2·2	11·1		$x - 12{\cdot}2$		$y - 4x + 40{\cdot}1$	
		$-1{\cdot}1$		$y - x + 3{\cdot}3$		$5x - 5y - 16{\cdot}3$
2·3	10		$y - 8{\cdot}9$		$x - 4y + 23{\cdot}8$	

x	y	Δ
		$y - 10$
2·4	y	
		$8{\cdot}2 - y$
2·5	8·2	
		$-0{\cdot}8$
2·6	7·4	

Taking $\Delta^5 y = 0$
$5x - 5y \cdot$
$-x + 10$
$\therefore \quad x = 12{\cdot}3$

Ex. 4-71. *Find the value of*
x : 2
y : 1
Sol.

x	y	Δ
2	1	
		4
3	5	
		8
4	13	
		$x - 13$
5	x	
		$61 - x$
6	61	
		64
7	125	

Assuming $\Delta^5 y = 0$, $x = 28$

Ex. 4-72. *The following t*
children born per mother. Find
years.

Age of mothers : 15
No. of children : (

Sol.

Age of mother in years
15 – 19
20 – 24
25 – 29
30 – 34
35 – 39
40 – 44

$\Delta^5 y_0 = 0$ or (
$\therefore \quad y_5 - 5y_4 + 10y_3 - 10$
$\therefore \quad 5{\cdot}8 - 5(5{\cdot}7) + 10y_3 -$
or $\quad y_3 = 4{\cdot}79.$

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5
		$y-10$		$27{\cdot}1-3y$		$10y-x-78{\cdot}1$
$2{\cdot}4$	y		$18{\cdot}2-2y$		$6y-54{\cdot}3$	
		$8{\cdot}2-y$		$3y-27{\cdot}2$		
$2{\cdot}5$	$8{\cdot}2$		$y-9$			
		$-0{\cdot}8$				
$2{\cdot}6$	$7{\cdot}4$					

Taking $\Delta^5 y = 0$

$$5x - 5y - 16{\cdot}3 = 0$$
$$-x + 10y - 78{\cdot}1 = 0$$

$\therefore$ $x = 12{\cdot}3$ and $y = 9{\cdot}04$.

Ex. 4-71. *Find the value of y for x = 5 from the set of values*

x :	2	3	4	6	7
y :	1	5	13	61	125

Sol.

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5
2	1					
		4				
3	5		4			
		8		$x-25$		
4	13		$x-21$		$120-4x$	
		$x-13$		$95-3x$		$10x-286$
5	x		$74-2x$		$6x-166$	
		$61-x$		$3x-71$		
6	61		$x+3$			
		64				
7	125					

Assuming $\Delta^5 y = 0$, $x = 28{\cdot}6$.

Ex. 4-72. *The following table gives the age of mother and the average number of children born per mother. Find the average number of children born per mother age 30–34 years.*

Age of mothers :	15 – 19	20 – 24	25 – 29	30 – 34	35 – 39	40 – 44
No. of children :	0·7	2·1	3·5	?	5·7	5·8

Sol.

Age of mother in years		*No. of children* y	
15 – 19	0	0·7	y_0
20 – 24	1	2·1	y_1
25 – 29	2	3·5	y_2
30 – 34	3	?	y_3
35 – 39	4	5·7	y_4
40 – 44	5	5·8	y_5

$$\Delta^5 y_0 = 0 \text{ or } (E-1)^5 y_0 = 0.$$

$\therefore$ $y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$

$\therefore$ $5{\cdot}8 - 5(5{\cdot}7) + 10y_3 - 10(3{\cdot}5) + 5(2{\cdot}1) - 0{\cdot}7 = 0.$

or $y_3 = 4{\cdot}79.$

Ex. 4-73. *Interpolate the missing figure in the following table of rice cultivation :*

Year	*Acres (in Millions)*	*Year*	*Acres (in Millions)*
1991	76·6	1996	?
1992	78·7	1997	50·6
1993	?	1998	77·6
1994	77·7	1999	78·6
1995	78·7		

Sol.

Year		*Acres*		*Year*		*Acres*	
1991	0	76·6	y_0	1996	5	?	y_5
1992	1	78·7	y_1	1997	6	50·6	y_6
1993	2	?	y_2	1998	7	77·6	y_7
1994	3	77·7	y_3	1999	8	78·6	y_8
1995	4	78·7	y_4				

As two missing terms are to be determined, two equations are needed. We take them to be

$$\Delta^7 y_0 = 0 \qquad ...(1)$$

and $$\Delta^7 y_1 = 0 \qquad ...(2)$$

Eqs. (1) and (2) give

$$y_7 - 7y_6 + 21y_5 - 35y_4 + 35y_3 - 21y_2 + 7y_1 - y_0 = 0$$

and $$y_8 - 7y_7 + 21y_6 - 35y_5 + 35y_4 - 21y_3 + 7y_2 - y_1 = 0$$

$\therefore \quad 21y_5 - 21y_2 = -162{\cdot}7$

or $\quad 35y_5 - 7y_2 = 1642{\cdot}1$

$\therefore \quad y_5 = 60{\cdot}58 \cong 60{\cdot}6$

and $\quad y_2 = 68{\cdot}3$

Ex. 4-74. *Estimate the production for the years* 1985 *and* 1985 *with the help of the following table :*

Year :	1970	1975	1980	1985	1990	1995	2000
Production in 000,000 units :	200	220	260	?	350	?	430

Sol.

x			
1970	0	200	y_0
1975	1	220	y_1
1980	2	260	y_2
1985	3	?	y_3
1990	4	350	y_4
1995	5	?	y_5
2000	6	430	y_6

We take $\quad \Delta^5 y_0 = 0$

and $\quad \Delta^5 y_1 = 0$

i.e., $\quad y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$

and $\quad y_6 - 5y_5 + 10y_4 - 10y_3 + 5y_2 - y_1 = 0$

$\therefore \quad y_5 + 10y_3 = 3450$

and $\quad 5y_5 + 10y_3 = 5010$

ing table of rice cultivation :

ear	Acres (in Millions)
996	?
997	50·6
998	77·6
999	78·6

	Acres	
5	?	y_5
6	50·6	y_6
7	77·6	y_7
8	78·6	y_8

ons are needed. We take them to

...(1)

...(2)

0

0

and 1985 with the help of the

1990	1995	2000
350	?	430

y_0
y_1
y_2
y_3
y_4
y_5
y_6

$\therefore \qquad y_5 = 390$ and $y_3 = 306.$

Ex. 4-75. *Interpolate U_2 from the following table :*

x	:	1	2	3	4	5
U_x	:	7	?	13	21	37

and explain why the value obtained is different from that obtained by putting $x = 2$ in the expression $2^x + 5$.

Sol.

x			
1	0	7	y_0
2	1	?	y_1
3	2	13	y_2
4	3	21	y_3
5	4	37	y_4

We take $\quad \Delta^4 y_0 = 0$

i.e., $\quad y_4 - 4y_3 + 6y_2 - 4y_1 + y_0 = 0$

$\therefore \qquad y_1 = 9{\cdot}5.$

EXERCISES

1. If (*i*) $f(x) = x^{n+1}$ show that

$$f(x_0, x_1, \ldots x_n) = x_0 + x_1 + \ldots + x_n$$

(*ii*) $f(x) = \dfrac{1}{x}$, show that

$$f(x_0, x_1, \ldots x_n) = \frac{(-1)^n}{x_0, x_1 \ldots x_n}$$

(*iii*) Show that *n*th order divided differences of a *n*th degree polynomial are constant and higher order divided differences are zero.

2. Show that

$$\Delta^n (ax^n + bx^{n-1}) = n!a \quad (h = 1)$$

3. Prove that

$$U_x - U_{x+1} + U_{x+2} + U_{x+2} + \ldots$$

$$= \frac{1}{2}\left[U_{x-\frac{1}{2}} - \frac{1}{8}\Delta^2 U_{x-\frac{3}{2}} = \frac{1 \cdot 3}{2!}\,\frac{1}{8^2}\,\Delta^2 U_{x-\frac{5}{2}} \ldots\right]$$

4. Show that

$$U_{2n} - {}^nc_1\, 2^1\, U_{2n-1} + {}^nc_2\, 2^2\, U_{2n-2} \ldots + (-1)^n\, 2^n\, U_n = (-1)^n \{c - 2\,an\}$$

where $\qquad U_x = ax^2 + bx + c$

and the interval of differencing is unity.

5. Find the Values of

(1) $2x^{(4)} + 3x^{(2)} + x^{(1)} - 7$ at $x = 5$. **[Ans. 298]**

(2) $2x^3 - 3x^2 + 3x - 10$ at $x = 5$. **[Ans. 180]**

6. Represent the following polynomials in the factorial notations.

(1) $x^4 - 12x^3 + 42x^2 - 30x + 9$ **[Ans. $x^{(4)} - 6x^{(3)} + 13x^{(2)} + x^{(1)} + 9$]**

(2) $x^4 - 3x^3 + 2x + 6.$ **[Ans. $x^{(4)} + 6x^{(3)} + 4x^{(2)} + 6$]**

7. Find a cubic function of x which has the values 1, – 3, – 1, 13 when $x = 1, 2, 3, 4$ respectively. **[Ans. $5 - 2x - 3x^2 + x^3$]**

8. Sum the series

(1) $2.3 + 3.6 + 4.11 + \ldots\ldots + (n+1)(n^2+2)$. $\left[\textbf{Ans. } \frac{n}{12}(3n^3+10n^2+21n+38\right]$

(2) $1^2.2^2 + 2^2.3^2 + 3^2.4^2 + \ldots$ upto n terms. $\left[\textbf{Ans. } \frac{n}{15}(3n^4+15n^3+25n^2+15n+2\right]$

9. Apply Lagrange's formula to find $f(5)$ and $f(6)$ given that

x	:	1	2	3	4	7
$f(x)$	:	2	4	8	16	128

and explain why the result differ from those obtained by completing the series of powers of 2. **[Ans.** 32·9 and 66·7]

10. If $P(x) = a_0 + a_1 \cos 2x + a_2 \sin 2x$ agrees with $f(x)$ when $x = x_0, x_1\, x_2$ then prove that

$$P(x) = \frac{\sin(x-x_1)\sin(x-x_2)}{\sin(x_0-x_1)\sin(x_0-x_2)}\, f(x_0) + \text{two similar terms.}$$

11. Use Newton's formula for interpolation to find the annual premium at the age of 33 from the table given below :

Age	24	28	32	36	40
Annual premium (in Rs.)	28·06	30·19	32·75	35·94	40·00

[Ans. 33·48]

12. From the following table estimate, by using Newton's formula, the premium payable at the age of 22 years :

Age (in yrs.)	20	25	30	35	40	45
Premium (in Rs.)	25	28	32	37	43·5	52·25

[Ans. 26·05]

13. Use Newton's formula to find the annual premium payable at the age of 26 years from the following table giving the annual premiums charged by an insurance company for a policy of Rs. 1,000 :

Age nex birthday :	20	25	30	35	40
Annual premium (in Rs.) :	23	26	30	35	42

[Ans. 26·73]

14. Using Newton's formula for interpolation estimate the population for the year 1965 :

Year	*Population*
1951	98,754
1961	132,285
1971	168,076
1981	195,690
1991	246,050

[Ans. 1,47,841]

15. From the following information find the number of students who obtained less than 45 Marks :

Marks	30–40	40–50	50–60	60–70	70–80
Frequency	31	42	51	35	31

[Ans. 48]

16. Determine the number of workers earning Rs. 124 or more but less than Rs. 125 from the following data :

Earnings less than Rs.	*No. of workers*	*Earnings less than Rs.*	*No. of workers*
120	276	135	918
125	599	140	966
130	804		

[Ans. 54]

17. From the following table es
Rs. 60 and Rs. 70.

Wage (in Rs.)	*No. (in t.*
Below 40	
40 — 60	
60 — 80	

18. The following table relates t
in a big manufacturing conc

Earnings per day	*No.*
up to Rs. 10	
up to Rs. 20	
up to Rs. 30	

Find the number of workers

19. Find $f(4)$ from table below

x	:	1	2
$f(x)$	:	2	4

20. Find (0·6538) using the foll

x	$f(x)$
0·62	0·6194114
0·63	0·6270463

21. Find log 324 using the follc

x	:	310	
$\log x$	:	2·491362	2

22. Use Stirling's formula to ol

x	:	0	0·
$f(x)$	:	0	0·19

23. If l_x represents the number
will permit. l_x for $x = 35, 4$

x	:	20	
l_x	:	512	4:

274]

24. Find $\sqrt{12516}$, using Gaus

x	:	12500
$\sqrt{x}$	:	111·803399

25. Use Stirling's formula to f
$u_{20} = 49225$,
$u_{35} = 45926$ and

26. Estimate the production o

year	:	1991	1'
In millions of bales	:	17·1	1

s. $\frac{n}{12}(3n^3+10n^2+21n+38]$

$(3n^4+15n^3+25n^2+15n+2]$

ıat

4 7

16 128

by completing the series of

[**Ans.** 32·9 and 66·7]

n $x = x_0, x_1\ x_2$ then prove that

ilar terms.

ual premium at the age of 33

40

4 40·00

[**Ans.** 33·48]

rmula, the premium payable

40 45

43·5 52·25

[**Ans.** 26·05]

able at the age of 26 years

ged by an insurance company

35 40

35 42

[**Ans.** 26·73]

opulation for the year 1965 :

[**Ans.** 1,47,841]

ents who obtained less than

)

[**Ans.** 48]

e but less than Rs. 125 from

s *No. of workers*

Rs.

918

966

[**Ans.** 54]

17. From the following table estimate the number of persons earning wages between Rs. 60 and Rs. 70.

Wage (in Rs.)	*No. of persons (in thousands)*	*Wage (in Rs.)*	*No. of persons (in thousands)*
Below 40	250	80 — 100	70
40 — 60	120	100 — 120	60
60 — 80	100		

[**Ans.** 54]

18. The following table relates to income earned per day by a certain number of workers in a big manufacturing concern :

Earnings per day	*No. of workers*	*Earnings per day*	*No. of workers*
up to Rs. 10	50	upto Rs. 40	500
up to Rs. 20	150	upto Rs. 50	700
up to Rs. 30	300	upto Rs. 60	800

Find the number of workers falling within the Rs. 25-35 earning group. [**Ans.** 178]

19. Find $f(4)$ from table below :

x :	1	2	3	5	6	7
$f(x)$:	2	4	8	32	64	128

[**Ans.** 16·1]

20. Find (0·6538) using the following data :

x	$f(x)$	x	$f(x)$	x	$f(x)$
0·62	0·6194114	0·64	0·6345857	0·67	0·6566275
0·63	0·6270463	0·65	0·6420292	0·68	0·6637820
		0·66	0·6493765		

[**Ans.** 0·6448325]

21. Find log 324 using the following data :

x :	310	320	330	340	350
$\log x$:	2·491362	2·505150	2·518514	2·531479	2·544068

[**Ans.** 2·4510545]

22. Use Stirling's formula to obtain $f(1{\cdot}22)$ from the following data :

x :	0	0·5	1·0	1·5	2·0	2·5	3·0
$f(x)$:	0	0·19146	0·34134	0·45319	0·47725	0·49379	0·49865

[**Ans.** 0·38871]

23. If l_x represents the number living at age x in a life table, find, as accurately as the data will permit. l_x for $x = 35$, 42 and 47. Given :

x :	20	30	40	50
l_x :	512	439	346	243

[**Ans.** 395, 326, 274]

24. Find $\sqrt{12516}$, using Gauss's backward formula, from the following data :

x :	12500	12510	12520	12530
$\sqrt{x}$:	111·803399	111·848111	111·892806	111·937483

[**Ans.** 111·874929]

25. Use Stirling's formula to find u_{28}, given that

$u_{20} = 49225$, $u_{25} = 48316$, $u_{30} = 47236$,
$u_{35} = 45926$ and $u_{40} = 44306$

26. Estimate the production of cotton in the year 1995 from the data given below :

year :	1991	1992	1993	1994	1995	1996
In millions of bales :	17·1	13·0	14·0	9·6	?	12·4

□□

5

Numerical Differentiation and Integration

5.1. Numerical Differentiation

It is the process of finding the derivatives of a function which may not be given in explicit mathematical form but for which a certain set of values are given. The procedure is to represent the function by an interpolation formula and then to differentiate this formula as many times as desired.

Rules of representing the function by an interpolation formula.

Argument Values		*Formula use*
Equidistant	*Newton's forward*	*(for differentiating near the beginning of the table)*
	Newton's backward	*(near the end)*
	Stirling's or Bessel's	*(near the middle)*
Non-equidistant	*Newton's divided difference* or *Lagrange's*	

Ex. 5-1. *Find the first and second order derivatives of the function tabulated below at the points* $x = 0, 0{\cdot}03$ *and* $0{\cdot}06$.

x :	0	0·01	0·02	0·03	0·04	0·05	0·06
$f(x)$:	0	0·0301	0·0604	0·0909	0·1216	0·1525	0·1836

Sol.

Difference Table

x	$f(x)$	Δ	Δ^2
0·00	0·0000		
		0·0301	
0·01	0·0301		0·0002
		0·0303	
0·02	0·0604		0·0002
		0·0305	
0·03	0·0909		0·0002
		0·0307	
0·04	0·1216		0·0002
		0·0309	
0·05	0·1525		0·0002
		0·0311	
0·06	0·1836		

(*i*) *At* $x = 0$

Since the differentiation is to be done near the beginning of the table, Newton's forward formula will be used. By the said formula,

$$f(x) = f(x_0) + U\Delta f(x_0) + \frac{U(U-1)}{2!}\Delta^2 f(x_0)$$

ıd Integration

on which may not be given in
'ues are given. The procedure is
ıen to differentiate this formula

on formula.

Formula use

(for differentiating near the beginning of the table)
(near the end)
(near the middle)

:e

the function tabulated below at

0·04	0·05	0·06
0·1216	0·1525	0·1836

Δ^2
0·0002
0·0002
0·0002
0·0002
0·0002

ȝ of the table, Newton's forward

$\frac{-1)}{} \Delta^2 f(x_0)$

(11 May 1692 - 05 Dec 1770)

Scottish Mathematician

Here $\quad U = \dfrac{x-0}{0\cdot 01} = 100x$

$$\therefore \quad f'(x) = \frac{df(x)}{dU}\,\frac{dU}{dx} = 100\left\{\Delta f(x_0) + \frac{2U-1}{2}\Delta^2 f(x_0)\right\}$$

$$= 100\ \{0\cdot 0301 + (U-0\cdot 5)\ (0\cdot 0002)\}$$

$$\therefore \quad f'(0) = 100\{0\cdot 0301 + (U-0\cdot 5)\ (0\cdot 0002)\}_{U=0}$$

$$= 100\ \{0\cdot 0301 - 0\cdot 0001\} = 3$$

Also $\quad f''(0) = \left\{\dfrac{df'(x)}{dU}\cdot\dfrac{dU}{dx}\right\}_{x=0} = (100)^2\ \{0\cdot 0002\} = 2$

(ii) At $x = 0\cdot 03$.

Since the differentiation is to be done near the middle of the table, any central difference formula can be used. By Stirling's formula,

$$f(x) = f(x_0) + U\mu\delta f(x_0) + \frac{U^2}{2!}\delta^2 f(x_0) + \frac{U(U^2-1)}{3!}\ \mu\delta^3 f(x_0) + \ldots$$

$$= f(x_0) + U\,\frac{E^{1/2}+E^{-1/2}}{2}\ \Delta f\left(x_0 - \frac{h}{2}\right) + \frac{U^2}{2!}\ \Delta^2 f(x_0-h)$$

$$+ \frac{U(U^2-1)}{3!}\,\frac{E^{1/2}+E^{-\frac{1}{2}}}{2}\ \Delta^3 f\left(x_0 - \frac{3h}{2}\right) + \ldots$$

$$= f(x_0) + \frac{U}{2}\ \{\Delta f(x_0) + \Delta f(x_0-h)\} + \frac{U^2}{2!}\ \Delta^2 f(x_0-h) + \ldots.$$

Here $\quad U = \dfrac{x-0\cdot 03}{0\cdot 01} = 100x - 3$

$$\therefore \quad f'(x) = 100\left\{\frac{df}{dU}\right\} = 100\left\{\frac{\Delta f(x_0)+\Delta f(x_0-h)}{2} + U\Delta^2 f(x_0-h)\right\}$$

$$= 100\left\{\frac{0\cdot 0307+0\cdot 0305}{2} + U(0\cdot 0002)\right\}$$

$$\therefore \quad f(0\cdot 03) = 100\ \{0\cdot 0306 + (0\cdot 0002)\ U\}_{U=0}$$

$$= 3\cdot 06$$

$$f''(0\cdot 03) = \left\{\frac{df'}{dU}\cdot\frac{dU}{dx}\right\}_{x=0\cdot 03} = (100)^2\ \{0\cdot 0002\} = 2.$$

(iii) At $x = 0\cdot 06$.

Since differentiation is to be done near the end of the table, Newton's backward formula will be used. By the said formula,

$$\therefore \quad f(x) = f(x_n) + U\nabla f(x_n) + \frac{U(U+1)}{2!}\ \nabla^2 f(x_n) + \ldots.$$

Here $\quad U = \dfrac{x-0\cdot 06}{0\cdot 01} = 100x - 6$

$$\therefore \quad f'(x) = 100\left\{\nabla f(x_n) + \frac{2U+1}{2}\ \nabla^2 f(x_n) + \ldots\right\}$$

$$= 100\ [0\cdot 0311 + (U+0\cdot 5)\ (0\cdot 0002)]$$

$$\therefore \quad f'(0\cdot06) = 100\,\{0\cdot0311 + (U+0\cdot5)\,(0\cdot0002)\}_{U=0}$$
$$= 100\,\{0\cdot0311 + 0\cdot0001\}$$
$$= 3\cdot12$$

$$f''(0\cdot06) = \left\{\frac{df'(x)}{dU}\,\frac{dU}{dx}\right\}_{x=0\cdot06} = (100)^2\,\{0\cdot0002\} = 2$$

Ex. 5-2. *Find the value of* $\cosh x = \frac{d}{dx}\{\sinh x\}$ *at* $x = 1\cdot52$ *from the following table :*

x	$\sinh x$	x	$\sinh x$
1·5	2·129279	1·8	2·942174
1·6	2·375568	1·9	3·268163
1·7	2·645632	2·0	3·626860

Sol.

Difference Table

x	$f(x) = \sinh x$	Δ	Δ^2	Δ^3	Δ^4	Δ^5
1·5	2·129279					
		0·246289				
1·6	2·375568		0·023775			
		0·270064		0·002703		
1·7	2·645632		0·026478		0·000266	
		0·296542		0·002969		0·000026
1·8	2·942174		0·029447		0.000292	
		0·325989		0·003261		
1·9	3·268163		0·032708			
		0·358697				
2·0	3·626860					

$$U = \frac{x-1\cdot5}{0\cdot1} = 10x - 15$$

For $x = 1\cdot52,\ U = 0\cdot2.$

Now from Newton's forward formula

$$f(x) = f(x_0) + U\Delta f(x_0) + \frac{U(U-1)}{2!}\,\Delta^2 f(x_0) + \frac{U(U-1)(U-2)}{3!}\,\Delta^3 f(x_0)$$
$$+ \frac{U(U-1)(U-2)(U-3)}{4!}\,\Delta^4 f(x_0)$$
$$+ \frac{U(U-1)(U-2)(U-3)(U-4)}{5!}\,\Delta^5 f(x_0)$$

(neglecting higher differences)

$$= f(x_0) + U\Delta f(x_0) + \frac{U^2-U}{2}\,\Delta^2 f(x_0) + \frac{U^2-3U^2+2U}{6}\,\Delta^3 f(x_0)$$
$$+ \frac{U^4-6U^3+11U^2-6U}{24}\,\Delta^4 f(x_0)$$
$$+ \frac{U^5-10U^4+35U^3-50U^2+24U}{120}\,\Delta^5 f(x_0)$$

$$\therefore \quad f'(x) = 10\left[\Delta f(x_0) + \frac{2U-1}{2}\,\Delta^2 f(x_0) + \frac{3U^2-6U+2}{6}\,\Delta^3 f(x_0)\right.$$

$+ \frac{4U^3-1}{}$

$+ \frac{5U^4-4}{}$

$f'(1\cdot52) = 10[0\cdot24628$

$+ \left(\frac{0\cdot46}{3}\right)(0\cdot$

$= 10\,[0\cdot2462$

$= 2\cdot395473.$

Ex. 5-3. *Assuming Newton*

(*i*) $f'(x_0) = \left\{\frac{df(x)}{dx}\right\}_{x =} =$

(*ii*) $f'(x_0) = \frac{1}{4}\{\Delta y_{-1} +$

where $y_0 = f(x_0)$ etc.

Sol. By the said formula,

$f(x) = y_0 + U\Delta y_0$

$\therefore \frac{df(x)}{dx} = \frac{1}{h}\frac{d\{f(x)\}}{dU}$

$= \frac{1}{h}\{\Delta y_0 +$

For $x = x_0,\ U =$

$\therefore f'(x_0) = \frac{1}{h}\{\Delta y_0 -$

To obtain (*ii*) put $x = x_1$ so t

$U = \frac{x_1-x_0}{h} =$

$\therefore f'(x_1) = \frac{1}{h}\{\Delta y_0 +$

Shifting the origin to x_1,

$f(x_0) = \frac{1}{h}\{\Delta y_{-1} +$

Ex. 5-4. *Assuming Newton*

$h^2 f''(x_0) = \Delta^2 y_0 - \Delta^3 y$

Sol. By the said formula

$f(x) = y_0 + U\Delta y_0$

Now $U = \frac{x-x_0}{h}$

$$+ \frac{4U^3 - 18U^2 + 22U - 6}{24} \Delta^4 f(x_0)$$

$$\left. + \frac{5U^4 - 40U^3 + 105U^2 - 100U + 24}{120} \Delta^5 f(x_0) \right]$$

$$f'(1{\cdot}52) = 10[0{\cdot}246289 + (-0{\cdot}3)(0{\cdot}023775)$$

$$\left. + \left(\frac{0\cdot46}{3}\right)(0\cdot002703) + \left(-\frac{0\cdot286}{3}\right)(0\cdot000266) + \left(\frac{0\cdot986}{15}\right)(0\cdot000026\right]$$

$$= 10\,[0{\cdot}246289 - 0{\cdot}0071325 + 0{\cdot}00041446 - 0{\cdot}00002536 + 0{\cdot}00000171]$$

$$= 2{\cdot}395473.$$

Ex. 5-3. *Assuming Newton's forward interpolation formula show that*

$$(i)\quad f'(x_0) = \left\{\frac{df(x)}{dx}\right\}_{x = x_0} = \frac{1}{h}\left\{\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0 + \ldots\right\}$$

$$(ii)\quad f'(x_0) = \frac{1}{4}\left\{\Delta y_{-1} + \frac{1}{2}\Delta^2 y_{-1} - \frac{1}{6}\Delta^3 y_{-1} + \ldots\right\}$$

where $y_0 = f(x_0)$ etc.

Sol. By the said formula,

$$f(x) = y_0 + U\Delta y_0 + \frac{U(U-1)}{2!}\Delta^2 y_0 + \frac{U(U-1)(U-2)}{3!}\Delta^3 y_0 + \ldots$$

$$\therefore \quad \frac{df(x)}{dx} = \frac{1}{h}\frac{d\{f(x)\}}{dU}$$

$$= \frac{1}{h}\left\{\Delta y_0 + \frac{2U-1}{2}\Delta^2 y_0 + \frac{3U^2 - 6U + 2}{6}\Delta^3 y_0 \ldots\right\}$$

For $\quad x = x_0,\ U = \dfrac{x_0 - x_0}{h} = 0$

$$\therefore \quad f'(x_0) = \frac{1}{h}\left\{\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 + \ldots\right\}$$

To obtain (*ii*) put $x = x_1$ so that

$$U = \frac{x_1 - x_0}{h} = \frac{h}{h} = 1$$

$$\therefore \quad f'(x_1) = \frac{1}{h}\left\{\Delta y_0 + \frac{1}{2}\Delta^2 y_0 - \frac{1}{6}\Delta^3 y_0 + \ldots\right\}$$

Shifting the origin to x_1,

$$f(x_0) = \frac{1}{h}\left\{\Delta y_{-1} + \frac{1}{2}\Delta^2 y_{-1} - \frac{1}{6}\Delta^3 y_{-1} + \ldots\right\}.$$

Ex. 5-4. *Assuming Newton's forward interpolation formula show that*

$$h^2 f''(x_0) = \Delta^2 y_0 - \Delta^3 y_0 + \ldots$$

Sol. By the said formula

$$f(x) = y_0 + U\Delta y_0 + \frac{U(U-1)}{2!}\Delta^2 y_0 + \frac{U(U-1)(U-2)}{3!}\Delta^3 y_0 + \ldots$$

Now $\quad U = \dfrac{x - x_0}{h}$

$$\therefore \quad x = x_0 + Uh$$

$$\therefore \quad f(x_0 + hU) = y_0 + U\Delta y_0 + \frac{U(U-1)}{2!}\Delta^2 y_0 + \frac{U(U-1)(U-2)}{3!}\Delta^3 y_0 + \ldots.. \qquad \ldots(1)$$

Now by Taylor's theorem

$$f(x_0 + hU) = f(x_0) + hUf'(x_0) + \frac{h^2U^2}{2!} f''(x_0) + \ldots.$$

$\therefore$ Equating co-efficients of U^2 in (1)

$$\frac{h^2}{2!} f''(x_0) = \frac{\Delta^2 y_0}{2!} - \frac{1}{2}\Delta^3 y_0 + \ldots..$$

$$\therefore \quad h^2 f''(x_0) = \Delta^2 y_0 - \Delta^3 y_0 + \ldots..$$

Note. Equating the co-efficients of various powers of U in (1) expressions for derivatives of various orders at $x = x_0$ can be obtained.

Ex. 5-5. *Assuming Newton's backward interpolation formula show that*

$$(i)\ hf'(x_0) = \nabla y_0 + \frac{1}{2}\nabla^2 y_0 + \frac{1}{3}\nabla^3 y_0 + \ldots.$$

$$(ii)\ hf'(x_0) = \nabla y_1 - \frac{1}{2}\nabla^2 y_1 - \frac{1}{6}\nabla^3 y_1 + \ldots.$$

Sol.

$$f(x) = y_0 + U\nabla y_0 + \frac{U(U+1)}{2!}\nabla^2 y_0 + \frac{U(U+1)(U+2)}{3!}\nabla^3 y_0 + \ldots.$$

(where the origin is at x_n)

$$\therefore \quad -hf'(x) = \frac{df(x)}{dU} = \nabla y_0 + \frac{2U+1}{2}\nabla^2 y_0 + \frac{3U^2+6U+2}{6}\nabla^3 y_0 \ldots$$

$$(i) \quad \therefore \quad hf'(x_0) = \nabla y_0 + \frac{1}{2}\nabla^2 y_0 - \frac{1}{3}\nabla^3 y_0 + \ldots.$$

(*ii*) Put $x = x_{-1}$ so that $U = -1$

$$\therefore \quad hf'(x_{-1}) = \nabla y_0 - \frac{1}{2}\nabla^2 y_0 - \frac{1}{6}\nabla^3 y_0 + \ldots.$$

Shifting the origin to x_{-1}

$$hf'(x_0) = \nabla y_1 - \frac{1}{2}\nabla^2 y_1 - \frac{1}{6}\nabla^3 y_1 + \ldots..$$

Ex. 5-6. *Show that*

$$(i)\ f'\left(x_0 + \frac{h}{2}\right) = \frac{1}{h}\left\{\Delta y_0 - \frac{1}{24}\Delta^3 y_0 + \ldots..\right\}$$

$$(ii)\ f'\left(x_0 - \frac{h}{2}\right) = \frac{1}{h}\left\{\nabla y_0 - \frac{1}{24}\nabla^3 y_0 + \ldots..\right\}.$$

Sol. $(i)\ f'\left(x_0 + \frac{h}{2}\right) = E^{1/2} Df(x_0)$

$$= (1+\Delta)^{1/2} Df(x_0)$$

Now $\quad D \equiv \frac{1}{h}\left\{\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 \ldots..\right\}$ **See Ex.** 5-3(*i*)

MATHEMATICAL STATISTICS

...(1)

$x_0) + ...$

(1) expressions for derivatives

mula show that

$$\frac{U(U+1)(U+2)}{3!}\nabla^3 y_0 + ...$$

$$\frac{3U^2+6U+2}{6}\nabla^3 y_0 ...$$

See Ex. 5-3(i)

$$
\begin{aligned}
\therefore \qquad f'\left(x_0+\frac{h}{2}\right) &= \frac{1}{h}(1+\Delta)^{1/2}\left(\Delta-\frac{1}{2}\Delta^2+\frac{1}{3}\Delta^3 ...\right) f(x_0) \\
&= \frac{1}{h}\left(1+\frac{1}{2}\Delta-\frac{1}{8}\Delta^2 ...\right)\left\{\Delta-\frac{1}{2}\Delta^2+\frac{1}{3}\Delta^3 ...\right\} f(x_0) \\
&= \frac{1}{h}\left\{\Delta-\frac{1}{24}\Delta^3 ...\right\} f(x_0) \\
&= \frac{1}{h}\left\{\Delta y_0-\frac{1}{24}\Delta^3 y_0 ...\right\}
\end{aligned}
$$

where $\qquad y_0 = f(x_0)$

Similarly (ii) can be proved.

Ex. 5-7. *Assuming Stirling's formula find the expressions for first six derivatives of f (x) at $x = x_0$.*

Sol. From the said formula

$$
\begin{aligned}
f(x) &= y_0 + U\mu\delta y_0 + \frac{U^2}{2!}\delta^2 y_0 + \frac{U(U^2-1)}{3!}\mu\delta^3 y_0 + ... \\
&+ \frac{U^2(U^2-1)...\{U^2-(r-1)^2\}}{2r!}\delta^{2r} y_0 \\
&+ \frac{U(U^2-1)...(U^2-r^2)}{(2r+1)!}\mu\delta^{2r+1} y_0 .... \\
&= y_0 + U\left\{\mu\delta y_0 - \frac{1}{6}\mu\delta^3 y_0 + \frac{1}{30}\mu\delta^5 y_0 - \frac{1}{140}\mu\delta^7 y_0 + ...\right\} \\
&+ \frac{U^2}{2!}\left\{\delta^2 y_0 - \frac{1}{12}\delta^4 y_0 + \frac{1}{90}\delta^6 y_0 - \frac{1}{560}\delta^8 y_0 + ...\right\} \\
&+ \frac{U^3}{3!}\left\{\mu\delta^3 y_0 - \frac{1}{4}\mu\delta^5 y_0 + \frac{7}{120}\mu\delta^7 y_0 ....\right\} \\
&+ \frac{U^4}{4!}\left\{\delta^4 y_0 - \frac{1}{6}\delta^6 y_0 + \frac{7}{240}\delta^8 y_0 ....\right\} \\
&+ \frac{U^5}{5!}\left\{\mu\delta^5 y_0 - \frac{1}{3}\mu\delta^7 y_0 ....\right\} \\
&+ \frac{U^6}{6!}\left\{\delta^6 y_0 - \frac{1}{4}\delta^8 y_0 + ...\right\} .....
\end{aligned}
$$

By Taylor's theorem,

$$f(x) = f(x_0+Uh) = f(x_0) + Uhf^{(1)}(x_0) + \frac{U^2h^2}{2!}f^{(2)}(x_0) + ...$$

Equating co-efficients of different powers of U

$$
\begin{aligned}
hf^{(1)}(x_0) &= \mu\delta y_0 - \frac{1}{6}\mu\delta^3 y_0 + \frac{1}{30}\mu\delta^5 y_0 - \frac{1}{140}\mu\delta^7 y_0 + ... \\
h^2 f^{(2)}(x_0) &= \delta^2 y_0 - \frac{1}{12}\delta^4 y_0 + \frac{1}{90}\delta^6 y_0 - \frac{1}{560}\delta^8 y_0 ..... \\
h^3 f^{(3)}(x_0) &= \mu\delta^3 y_0 - \frac{1}{4}\mu\delta^5 y_0 + \frac{7}{120}\mu\delta^7 y_0 + ...
\end{aligned}
$$

$$h^4 f^{(4)}(x_0) = \delta^4 y_0 - \frac{1}{6}\delta^6 y_0 + \frac{7}{240}\delta^8 y_0 \;....$$

$$h^5 f^{(5)}(x_0) = \mu\delta^5 y_0 - \frac{1}{3}\mu\delta^7 y_0 \;....$$

$$h^6 f^{(6)}(x_0) = \delta^6 y_0 - \frac{1}{4}\delta^8 y_0 \;.....$$

Ex. 5-8. *Show that*

$$\frac{dy_x}{dx} = \frac{2}{3}(y_{x+1} - y_{x-1}) - \frac{1}{12}(y_{x+2} - y_{x-2}).$$

Sol. From Ex. 5-7 (taking $h = 1$)

$$\left(\frac{dy_x}{dx}\right)_{at\, x = x_0} = \mu\delta y_0 - \frac{1}{6}\mu\delta^3 y_0 \text{ (neglecting higher differences)}$$

$$= \frac{1}{2}(E^{1/2} + E^{-1/2})(E^{1/2} - E^{-1/2})y_0$$

$$- \frac{1}{12}(E^{1/2} + E^{-1/2})(E^{1/2} - E^{-1/2})^3 y_0$$

$$= \frac{1}{2}(E - E^{-1})y_0 - \frac{1}{12}(E + E^{-1} - 2)(E - E^{-1})y_0$$

$$= \frac{1}{2}(y_1 + y_{-1}) - \frac{1}{12}\{E^2 - E^{-2} - 2(E - E^{-1})\}y_0$$

$$= \frac{1}{2}(y_1 - y_{-1}) - \frac{1}{12}(y_2 - y_{-2}) + \frac{1}{6}(y_1 - y_{-1})$$

$$= \frac{2}{3}(y_1 - y_{-1}) - \frac{1}{12}(y_2 - y_{-2})$$

Shifting the origin to x

$$\frac{dy_x}{dx} = \frac{2}{3}(y_{x+1} - y_{x-1}) - \frac{1}{12}(y_{x+2} - y_{x-2}).$$

Ex. 5-9. *Starting with Bessel's formula obtain the expressions of first four derivatives of $f(x)$ at $x = x_0$.*

Sol. From the said formula

$$f(x) = \mu y_{1/2} + \left(U - \frac{1}{2}\right)\delta y_{1/2} + \frac{U(U-1)}{2!}\mu\delta^2 y_{1/2}$$

$$+ \left(U - \frac{1}{2}\right)\frac{U(U-1)}{3!}\delta^3 y_{1/2} + ...$$

$$+ \frac{(U+r-1)(U+r-2)....(U-r)}{2r!}\mu\delta^{2r} y_{1/2}$$

$$+ \frac{\left(U - \frac{1}{2}\right)(U+r-1)(U+r-2)....(U-r)}{(2r+1)!}\delta^{2r+1} y_{1/2} + ...$$

Now $\mu y_{1/2} + \left(U - \frac{1}{2}\right)\delta y_{1/2} = \frac{y_1 + y_0}{2} + U\delta y_{1/2} - \frac{1}{2}(y_1 - y_0)$

$$= y_0 + U\delta y_{1/2}$$

$\therefore \quad f(x) =$

Friedrich Wilhelm Bessel (22 Jul 1784 - 17 Mar 1846) German Astronomer, Mathematician and Geodesist

$=$

$\therefore$ As in Ex. 5-7.

$hf^{(1)}(x_0) =$

$hf^{(2)}(x_0) =$

$hf^{(3)}(x_0) =$

$h^4 f^{(4)}(x_0) =$

Ex. 5-10. *Using Bessel's*

$\frac{dy_x}{dx} =$

Sol. From the said formu

$y_x =$

$=$

$$
\begin{aligned}
\therefore \quad f(x) &= y_0 + U\delta y_{1/2} + \frac{U(U-1)}{2!}\,\mu\delta^2 y_{1/2} \\
&+ \left(U - \frac{1}{2}\right)\frac{U(U-1)}{3!}\,\delta^3 y_{1/2} \\
&+ \ldots + \frac{(U+r-1)(U+r-2)\ldots(U-r)}{2r!}\,\mu\delta^{2r} y_{1/2} \\
&+ \frac{\left(U-\frac{1}{2}\right)(U+r-1)(U+r-2)\ldots(U-r)}{(2r+1)!}\,\delta^{2r+1} y_{1/2} + \ldots
\end{aligned}
$$

Friedrich Wilhelm Bessel (22 Jul 1784 – 17 Mar 1846) German Astronomer, Mathematician and Geodesist

$$
\begin{aligned}
&= y_0 + U\left\{\delta y_{1/2} - \frac{1}{2}\,\mu\delta^2 y_{1/2} + \frac{1}{12}\delta^3 y_{1/2} + \frac{1}{12}\mu\delta^4 y_{1/2} - \frac{1}{120}\delta^5 y_{1/2} + \ldots\right\} \\
&+ \frac{U^2}{2!}\left\{\mu\delta^2 y_{1/2} - \frac{1}{2}\,\delta^3 y_{1/2} - \frac{1}{12}\mu\delta^4 y_{1/2} + \frac{1}{24}\delta^5 y_{1/2} \ldots\right\} \\
&+ \frac{U^3}{3!}\left\{\delta^3 y_{1/2} - \frac{1}{2}\mu\delta^4 y_{1/2} \ldots\right\} \\
&+ \frac{U^4}{4!}\left(\mu\delta^4 y_{1/2} - \frac{1}{2}\delta^5 y_{1/2}\right) + \ldots
\end{aligned}
$$

$\therefore$ As in Ex. 5-7.

$$
\begin{aligned}
hf^{(1)}(x_0) &= \delta y_{1/2} - \frac{1}{2}\,\mu\delta^2 y_{1/2} + \frac{1}{12}\,\delta^3 y_{1/2} + \frac{1}{12}\,\mu\delta^4 y_{1/2} - \frac{1}{120}\,\delta^5 y_{1/2} \ldots\ldots \\
hf^{(2)}(x_0) &= \mu\delta^2 y_{1/2} - \frac{1}{2}\,\delta^3 y_{1/2} - \frac{1}{12}\,\mu\delta^4 y_{1/2} + \frac{1}{24}\,\delta^5 y_{1/2} \ldots\ldots \\
hf^{(3)}(x_0) &= \delta^3 y_{1/2} - \frac{1}{2}\,\mu\delta^4 y_{1/2} \ldots\ldots \\
h^4 f^{(4)}(x_0) &= \mu\delta^4 y_{1/2} - \frac{1}{2}\,\delta^5 y_{1/2} \ldots\ldots
\end{aligned}
$$

Ex. 5-10. *Using Bessel's formula show that*

$$
\frac{dy_x}{dx} = \Delta y_{x-\frac{1}{2}} - \frac{1}{24}\Delta^3 y_{x-\frac{3}{2}} + \ldots
$$

Sol. From the said formula,

$$
\begin{aligned}
y_x &= \mu y_{1/2} + \left(U - \frac{1}{2}\right)\delta y_{1/2} + \frac{U(U-1)}{2!}\,\mu\delta^2 y_{1/2} \\
&+ \left(U - \frac{1}{2}\right)\frac{U(U-1)}{3!}\,\delta^3 y_{1/2} + \ldots \\
&= \frac{y_1 + y_0}{2} + \left(U - \frac{1}{2}\right)\Delta y_0 + \frac{U(U-1)}{2!}\cdot\frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2} \\
&+ \left(U - \frac{1}{2}\right)\frac{U(U-1)}{3!}\,\Delta^3 y_{-1} + \ldots
\end{aligned}
$$

Here $\quad U = x - x_0 \qquad (\because h = 1)$

$\therefore$ Change x to $x + \frac{1}{2}$. Then U changes to $U + \frac{1}{2}$.

$$\therefore \quad y_{x+1/2} = \frac{y_1 + y_0}{2} + U\Delta y_0 + \frac{\left(U+\frac{1}{2}\right)\left(U-\frac{1}{2}\right)}{2!} \cdot \frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2}$$

$$+ \frac{U\left(U+\frac{1}{2}\right)\left(U-\frac{1}{2}\right)}{3!} \Delta^3 y_{-1} + \ldots.$$

$$\therefore \quad \frac{dy_{x+\frac{1}{2}}}{dx} = \Delta y_0 + \frac{2U}{2!} \frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2} + \frac{3U^2 - \frac{1}{4}}{3!} \Delta^3 y_{-1} + \ldots.$$

Put $x = x_0$ so that $U = 0$

$$\therefore \quad \left(\frac{dy_{x+\frac{1}{2}}}{dx}\right)_{x=x_0} = \Delta y_0 - \frac{1}{24} \Delta^3 y_{-1} + \ldots..$$

Shifting the origin from x_0 to $x - \frac{1}{2}$

$$\frac{dy_x}{dx} = \Delta y_{x-\frac{1}{2}} - \frac{1}{24} \Delta^3 y_{x-\frac{3}{2}} + \ldots.$$

5.2. Numerical Integration

It is the process of computing the value of a definite integral from a set of numerical values of the integrand. When the function to be integrated is of single variable the process is called **Mechanical Quadrature.** *The procedure is to represent the integrand by an interpolation formula and then to integrate this formula between the desired limits.*

Quadrature Formulae.

(1) Trapezoidal rule.

$$\int_{x_0}^{x_0+nh} y\,dx = h\left\{\frac{y_0 + y_n}{2} + (y_1 + y_2 + \ldots.. + y_{n-1})\right\}$$

(2) Simpson's one-third rule.

$$\int_{x_0}^{x_0+nh} y\,dx = \frac{h}{3}\left[(y_0 + y_n) + 4(y_1 + y_3 + \ldots. + y_{n-1}) + 2(y_2 + y_4 + \ldots. + y_{n-2})\right]$$

(*Can be applied only when 'n' is even*)

(3) Simpson's three-eighth rule.

$$\int_{x_0}^{x_0+nh} y\,dx = \frac{3}{8} h\left[(y_0 + y_n) + 3\{(y_1 + y_2) + (y_4 + y_5) \ldots. + (y_{n-2} + y_{n-1})\}\right.$$

$$\left. + 2(y_3 + y_6 + \ldots. + y_{n-3})\right]$$

(*Can be applied only when 'n' is a multiple of '3'*)

Thomas (20 Aug 17

Britex

(4) Weddle's rule.

$$\int_{x_0}^{x_0+nh} y\,dx =$$

(*Can be applied only wh*

Ex. 5-11. *Derive gener*

from it.

(1) *The Trapezoidal rule*

(2) *Simpson's one-third*

(3) *Simpson's three-eigh*

(4) *Weddle's rule.*

Sol. Let $y = f(x)$ be the

$I =$

Divide the range 'a' to '

$x_0 =$

Let the values of y at the

The method is to repres

integrate this formula betwe

formula of forward differenc

$f(x) =$

where $\quad U =$

$\therefore \quad I =$

$=$

$\frac{1}{2}\Big) \cdot \frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2}$

...

$-\frac{1}{4}\ \Delta^3 y_{-1} + \ldots$

gral from a set of numerical
of single variable the process
present the integrand by an
een the desired limits.

$-1)\Big\}$

$y_{n-1}) + 2(y_2 + y_4 + \ldots + y_{n-2})]$

$y_4 + y_5) \ldots + (y_{n-2} + y_{n-1})\}$

Thomas Simpson
(20 Aug 1710 – 14 May 1761)

British Mathematician

(4) Weddle's rule.

$$\int_{x_0}^{x_0+nh} y\,dx = \frac{3h}{10}[(y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6)$$
$$+ (y_6 + 5y_7 + y_8 + 6y_9 + y_{10} + 5y_{11} + y_{12}) + \ldots$$
$$+ (y_{n-6} + 5y_{n-5} + y_{n-4} + 6y_{n-3} + y_{n-2} + 5y_{n-1} + y_n)]$$

(*Can be applied only when 'n' is a multiple of* '6'.

Ex. 5-11. *Derive general quadrature formula for equidistant ordinates and deduce from it.*

(1) *The Trapezoidal rule.*
(2) *Simpson's one-third rule.*
(3) *Simpson's three-eighth rule.*
(4) *Weddle's rule.*

Sol. Let $y = f(x)$ be the function and it is required to evaluate

$$I = \int_a^b f(x)\,dx.$$

Divide the range 'a' to 'b' into n-equal parts and let the points of division be

$$x_0 = a, x_1 = x_0 + h, \ldots\ldots x_n = x_0 + nh = b.$$

Let the values of y at these points of division be $y_0, y_1 \ldots y_n$.

The method is to represent the integrand $f(x)$ by an interpolation formula and then to integrate this formula between the desired limits. Thus representing $f(x)$ by Newton's formula of forward differences.

$$f(x) = y_0 + U\Delta y_0 + \frac{U(U-1)}{2!}\Delta^2 y_0 + \frac{U(U-1)(U-2)}{3!}\Delta^3 y_0 + \ldots$$

where
$$U = \frac{x - x_0}{h}$$

$$\therefore \quad I = \int_a^b f(x)dx = \int_{x_0}^{x_0+nh} f(x)dx$$

$$= h\int_0^n \Big\{y_0 + U\Delta y_0 + \frac{U(U-1)}{2!}\Delta^2 y_0 + \frac{U(U-1)(U-2)}{3!}\Delta^3 y_0$$
$$+ \frac{U(U-1)(U-2)(U-3)}{4!}\Delta^4 y_0$$
$$+ \frac{U(U-1)(U-2)(U-3)(U-4)}{5!}\Delta^5 y_0$$
$$+ \frac{U(U-1)(U-2)(U-3)(U-4)(U-5)}{6!}\Delta^6 y_0 + \ldots\Big\} dU$$

$$= h\Bigg[ny_0 + \frac{n^2}{2}\Delta y_0 + \left(\frac{n^3}{3} - \frac{n^2}{2}\right)\frac{\Delta^2 y_0}{2!} + \left(\frac{n^4}{4} - n^3 + n^2\right)\frac{\Delta^3 y_0}{3!}$$
$$+ \left(\frac{n^5}{5} - \frac{3}{2}n^4 + \frac{11}{3}n^3 - 3n^2\right)\frac{\Delta^4 y_0}{4!} + \left(\frac{n^6}{6} - 2n^5 + \frac{35}{4}n^4\right.$$

$$-\frac{50}{3}n^3+12n^2\Big)\frac{\Delta^5 y_0}{5!}+\left(\frac{n^7}{7}-\frac{15n^6}{6}+17n^5-\frac{225n^4}{4}\right.$$

$$\left.+\frac{274}{3}n^3-60n^2\right)\frac{\Delta^6 y_0}{6!}+\ldots\ldots\Bigg] \qquad \ldots(A)$$

This formula is general quadrature formula.

(1) **Trapezoidal rule.**

Putting $n = 1$ in (A)

$$I_1 = \int_{x_0}^{x_1} f(x)dx = h\left\{y_0+\frac{1}{2}\Delta y_0\right\}.$$

Second and higher differences have been neglected as, since the interval of integration extends from x_0 to $x_1 = x_0 + h$, there are only two values and with these there can be no differences higher than the one.

$$\therefore \qquad I_1 = h\left\{y_0+\frac{y_1-y_0}{2}\right\} = \frac{h}{2}(y_0+y_1)$$

Similarly,
$$I_2 = \int_{x_1}^{x_2} f(x)dx = \frac{h}{2}(y_1+y_2)$$

....

$$I_n = \int_{x_{n-1}}^{x_n} f(x)dx = \frac{h}{2}(y_{n-1}+y_n)$$

$$\therefore \qquad I = \int_{x_0}^{x_0+nh} f(x)dx = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \ldots\ldots + \int_{x_{n-1}}^{x_n} f(x)dx$$

$$= I_1 + I_2 + \ldots. I_n$$

$$= \frac{h}{2}\{y_0 + 2(y_1+y_2+\ldots.+y_{n-1})+y_n\}$$

$$= h\left\{\frac{(y_0+y_n)}{2}+(y_1+y_2+\ldots\ldots+y_{n-1})\right\}.$$

(2) **Simpson's one-third rule.**

Putting $n = 2$ in (A) and neglecting third and higher differences.

$$I_1 = \int_{x_0}^{x_2} f(x)dx = h\left\{2y_0+2\Delta y_0+\frac{1}{3}\Delta^2 y_0\right\}$$

$$= h\left\{2y_0+2(y_1-y_0)+\frac{1}{3}(E-1)^2 y_0\right\}$$

$$= h\left\{2y_0+2(y_1-y_0)+\frac{1}{3}(y_2-2y_1+y_0)\right\}$$

$$= \frac{h}{3}\{y_0+4y_1+y_2\}$$

Similarly
$$I_2 = \int_{x_2}^{x_4} f(x)dx = \frac{h}{3}\{y_2+4y_3+y_4\}$$

....

$$I_{n/2} = \int_{x_{n-2}}^{x_n} f(x)dx = \frac{h}{3}\ \{y_{n-2} + 4y_{n-1} + y_n\} \quad \text{(assuming } n \text{ to be even)}$$

$$\therefore \quad I = \int_{x_0}^{x_n} f(x)\,dx = I_1 + I_2 + ... + I_{n/2}$$

$$= \frac{h}{3}\ [(y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + ... + (y_{n-2} + 4y_{n-1} + y_n)]$$

$$= \frac{h}{3}\ [(y_0 + y_n) + 4(y_1 + y_3 + ... + y_{n-1}) + 2(y_2 + y_4 + ... + y_{n-2})]$$

(3) Simpson's three-eighth rule.

Putting $n = 3$ in (A) and neglecting all differences above the third.

$$I_1 = \int_{x_0}^{x_3} f(x)\,dx = h\left\{3y_0 + \frac{9}{2}\Delta y_0 + \frac{9}{4}\Delta^2 y_0 + \frac{3}{8}\Delta^3 y_0\right\}$$

$$= h\left\{3y_0 + \frac{9}{2}(y_1 - y_0) + \frac{9}{4}(E-1)^2\, y_0 + \frac{3}{8}(E-1)^3\, y_0\right\}$$

$$= h\left\{3y_0 + \frac{9}{2}(y_1 - y_0) + \frac{9}{4}(E^2 - 2E + 1)y_0 + \frac{3}{8}(E^3 - 3E^2 + 3E - 1)y_0\right\}$$

$$= \frac{3}{8}h\{y_0 + 3y_1 + 3y_2 + y_3\}$$

Similarly,

$$I_2 = \int_{x_3}^{x_6} f(x)\,dx = \frac{3}{8}\ h\ \{y_3 + 3y_4 + 3y_5 + y_6\}$$

....

$$I_{n/3} = \int_{x_{n-3}}^{x_n} f(x)dx = \frac{3}{8}\ h\ \{y_{n-3} + 3y_{n-2} + 3y_{n-1} + y_n\}$$

(assuming n to be a multiple of 3)

$$\therefore \quad I = \int_{x_0}^{x_n} f(x)\,dx = I_1 + I_2 + ... I_{n/3}$$

$$= \frac{3}{8}\ h[y_0 + y_n) + 3\{(y_1 + y_2) + (y_4 + y_5) + ... + (y_{n-2} + y_{n-1})\}$$
$$+ 2(y_3 + y_6 + ... + y_{n-3})]$$

(4) Weddle's rule.

Putting $n = 6$ in (A) and neglecting differences above sixth.

$$I_1 = \int_{x_0}^{x_6} f(x)dx = h\Bigg[6y_0 + 18\Delta y_0 + 27\Delta^2 y_0 + 24\Delta^3 y_0 + \frac{123}{10}\Delta^4\, y_0$$
$$+ \frac{33}{10}\Delta^5 y_0 + \frac{41}{140}\Delta^6 y_0\Bigg]$$

$$= h\Big[6y_0 + 18(E-1)y_0 + 27(E-1)^2\, y_0 + 24(E-1)^3 y_0$$

$$+\frac{123}{10}\,(E-1)^4 y_0+\frac{33}{10}(E-1)^5 y_0+\frac{3}{10}(E-1)^6 y_0\Big]-\frac{1}{140}h\Delta^6 y_0$$

$$=\frac{3h}{10}\{y_0+5y_1+y_2+6y_3+y_4+5y_5+y_6\}-\frac{1}{140}\,h\Delta^6 y_0$$

Choosing 'h' *s.t.* the sixth differences are small, the last term can be neglected.

Then $$I_1 = \frac{3h}{10}\{y_0+5y_1+y_2+6y_3+y_4+5y_5+y_6\}$$

Similarly,

$$I_2 = \int_{x_6}^{x_{12}} f(x)dx=\frac{3h}{10}\,(y_6+5y_7+y_8+6y_9+y_{10}+5y_{11}+y_{12})$$

....

$$I_{\frac{n}{6}} = \int_{x_{n-6}}^{x_n} f(x)dx=\frac{3h}{10}\,\{y_{n-6}+5y_{n-5}+y_{n-4}+6y_{n-3}+y_{n-2}+5y_{n-1}+y_n\}$$

(assuming n to be a multiple of 6)

$$\therefore \quad I = \int_{x_0}^{x_n} f(x)\,dx = I_1+I_2+.....+I_{n/6}$$

$$=\frac{3h}{10}\,[(y_0+5y_1+y_2+6y_3+y_4+5y_5+y_6)+(y_6+5y_7+y_8+6y_9+y_{10}$$
$$+5y_{11}+y_{12})+.....+(y_{n-6}+5y_{n-5}+y_{n-4}+6y_{n-3}+y_{n-2}+5y_{n-1}+y_n)]$$

Note. Since in 'Trapezoidal Rule' second and higher difference are neglected, y is assumed to be linear. Similarly, in 'Simpson one-third rule', 'Simpson's three-eighth rule' and 'Weddle's rule' y is assumed to be polynomials of degree second, third and sixth respectively.

Ex. 5-12. *Evaluate* $I=\int_{1\cdot 0}^{1\cdot 3}\sqrt{x}\,dx$

by (1) *Simpson's rule.*
(2) *Trapezoidal rule.*

Sol. Take $h = 0\cdot1$

x :	1·0	1·1	1·2	1·3
$y=\sqrt{x}$:	1·000000	1·048809	1·095445	1·140175

(1) By Simpson's three-eighth rule

$$I = \frac{3}{8}h[(y_0+y_3)+3\{y_1+y_2\}]$$
$$=\frac{3}{8}\,(0\cdot1)\,[2\cdot1401175+6\cdot432762]$$
$$= 0\cdot321485$$

(2) By Trapezoidal rule.

$$I = h\left[\frac{y_0+y_3}{2}+(y_1+y_2)\right]$$
$$=(0\cdot1)\,[1\cdot0700875+2\cdot144254]$$
$$=0\cdot32143415\simeq 0\cdot321434.$$

Ex. 5-13. *Evaluate* $I =$

by (1) *Trapezoidal rul*
(2) *Simpson's rule.*

Sol. Take $h = \frac{\pi}{20}$

x	$y = \cos$
0	1·0000
$\frac{\pi}{20}$	0·9876
$\frac{2\pi}{20}$	0·9510
$\frac{3\pi}{20}$	0·8910
$\frac{4\pi}{20}$	0·8090
$\frac{5\pi}{20}$	0·7071

(1) By Trapezoidal rule,

$I = h\Big[\frac{y_0}{}$
$= \frac{\pi}{20}\,[1$
$= 0\cdot9983$

(2) By Simpson's one-th

$I = \frac{h}{3}\,[(y$
$= \frac{\pi}{20}\,[1$
$= 1\cdot0004$

Ex. 5-14. *Evaluate* $I =$

by (1) *Weddle's rule.*
(2) *Simpson's one*
(3) *Simpson's thr*

Sol. Divide the range c

x	$f(x) =$
0·0	1·000
0·5	0·800
1·0	0·500

Ex. 5-13. *Evaluate* $I = \int_0^{\pi/2} \cos x \, dx$

by (1) *Trapezoidal rule.*
(2) *Simpson's rule.*

Sol. Take $h = \dfrac{\pi}{20}$

x	$y = \cos x$	x	$y = \cos x$
0	1·000000	$\frac{6\pi}{20}$	0·587785
$\frac{\pi}{20}$	0·987688	$\frac{7\pi}{20}$	0·453990
$\frac{2\pi}{20}$	0·951057	$\frac{8\pi}{20}$	0·309017
$\frac{3\pi}{20}$	0·891007	$\frac{9\pi}{20}$	0·156434
$\frac{4\pi}{20}$	0·809017	$\frac{10\pi}{20}$	0·000000
$\frac{5\pi}{20}$	0·707107		

(1) By Trapezoidal rule,

$$I = h\left[\frac{y_0 + y_{10}}{2} + (y_1 + y_2 + \ldots + y_9)\right]$$

$$= \frac{\pi}{20}[10 \cdot 500000 + 5 \cdot 853102]$$

$$= 0 \cdot 9983446 \simeq 0 \cdot 998345$$

(2) By Simpson's one-third rule.

$$I = \frac{h}{3}[(y_0 + y_{10}) + 4(y_1 + y_3 + \ldots + y_9) + 2(y_2 + y_4 + \ldots + y_8)]$$

$$= \frac{\pi}{20}[1 \cdot 000000 + 12 \cdot 784904 + 5 \cdot 313752]$$

$$= 1 \cdot 000406.$$

Ex. 5-14. *Evaluate* $I = \int_0^6 \frac{dx}{1+x^2}$

by (1) *Weddle's rule.*
(2) *Simpson's one-third rule.*
(3) *Simpson's three-eighths rule.*

Sol. Divide the range of integration into twelve equal parts by taking $h = 0 \cdot 5$

x	$f(x) = \frac{1}{1+x^2}$	x	$f(x) = \frac{1}{1+x^2}$
0·0	1·00000	3·5	0·0754717
0·5	0·800000	4·0	0·0588235
1·0	0·500000	4·5	0·0470588

(*Contd.*)

x	$f(x)=\dfrac{1}{1+x^2}$	x	$f(x)=\dfrac{1}{1+x^2}$
1·5	0·307692	5·0	0·0384615
2·0	0·200000	5·5	0·0320000
2·5	0·137931	6·0	0·0270270
3·0	0·100000		

(1) By Weddle's rule.

$$\begin{aligned} I &= \frac{3}{10}(0\cdot5)\,[y_0+5y_1+y_2+6y_3+y_4+5y_5+y_6] \\ &\quad + (y_6+5y_7+y_8+6y_9+y_{10}+5y_{11}+y_{12})] \\ &= (0\cdot15)\,[8\cdot335807+1\cdot0440233] \\ &= 1\cdot406974545 \simeq 1\cdot407 \end{aligned}$$

(2) By Simpson's one-third rule.

$$\begin{aligned} I &= \frac{0\cdot5}{3}\,[(y_0+y_{12})+4(y_1+y_3+\ldots\ldots+y_{11})+2(y_2+y_4+\ldots.\,y_{10})] \\ &= \frac{0\cdot5}{3}\,[1\cdot0270270+5\cdot6006140+1\cdot7945700] \\ &= 1\cdot4037018 \simeq 1\cdot404. \end{aligned}$$

(3) By Simpson's three-eighth rule.

$$\begin{aligned} I &= \frac{3}{8}(0\cdot5)\,[(y_0+y_{12})+3(y_1+y_2)+(y_4+y_5)+(y_7+y_8)+(y_{10}+y_{11})] \\ &\quad +2(y_3+y_6+y_9)] \\ &= \frac{3}{8}(0\cdot5)\,[1\cdot0270270+5\cdot5280631+0\cdot9095016] \\ &= 1\cdot39961 \simeq 1\cdot400. \end{aligned}$$

Ex. 5-15. *Compute by Simpson's rule the value of the integral*

$$I = \int_{200}^{1000} \frac{dx}{\log_{10} x}$$

taking eight subintervals.

Sol. Here $h = 100$

x	$\log_{10} x$	$y=(\log_{10} x)^{-1}$	x	$\log_{10} x$	$y=(\log_{10} x)^{-1}$
200	2·3010	0·434594	700	2·8451	0·3514815
300	2·4771	0·403698	800	2·9031	0·344459
400	2·6021	0·384305	900	2·9542	0·338501
500	2·6990	0·370508	1000	3·0000	0·333333
600	2·7782	0·359945			

By Simpson's one-third rule,

$$\begin{aligned} I &= \frac{100}{3}\,[(y_0+y_8)+4(y_1+y_3+y_5+y_7)+2(y_2+y_4+y_6)] \\ &= \frac{100}{3}\,[0\cdot767927+5\cdot856754+2\cdot177418] \\ &= 293\cdot4033 \simeq 293\cdot4. \end{aligned}$$

Ex. 5-16. *Compute* $\log_e 2$
evaluate

$$\int_0^3 \frac{dx}{1+x}.$$

Sol. Divide the range of int

x	$f(x)=\dfrac{1}{1+}$
0·0	1·000000
0·5	0·666667
1·0	0·500000
1·5	0·400000

By Weddle's rule,

$$I = \int_0^3 \frac{dx}{1+x} =$$

$$+6\,(0\cdot400$$

Also $\quad I = \int_0^3 \frac{dx}{1+x} =$

$\therefore \quad 2\log_e 2 = 1\cdot386785$

$\therefore \quad \log_e 2 = 0\cdot693392$

Ex. 5-17. *Applying 'Simpso*

$$\int_0^1 \frac{dx}{1+x}$$

correct to three decimal places.

Sol. Divide the range of int

x	$f(x)=\dfrac{1}{1+}$
0·0	1·000000
0·1	0·909091
0·2	0·833333
0·3	0·769231
0·4	0·714286
0·5	0·666667

$$\begin{aligned} \therefore \quad I &= \int_0^1 \frac{dx}{1+x} = \\ &= \frac{0\cdot1}{3}\,\{1\cdot50 \\ &= 0\cdot69315 \\ &= 0\cdot693. \end{aligned}$$

Ex. 5-18. *Evaluate* $I = \int$

$f(x)=\dfrac{1}{1+x^2}$
0·0384615
0·0320000
0·0270270

$+ y_6]$

$)]$

$- 2(y_2 + y_4 + ... y_{10})]$

700]

$) + (y_7 + y_8) + (y_{10} + y_{11})]$

5016]

gral

$_{10}x$	$y=(\log_{10} x)^{-1}$
451	0·3514815
031	0·344459
542	0·338501
000	0·333333

$_2 + y_4 + y_6)]$

Ex. 5-16. *Compute $\log_e 2$ using a suitable quadrature formula with 7 ordinates to evaluate*

$$\int_0^3 \frac{dx}{1+x}.$$

Sol. Divide the range of integration into six parts by taking $h = 0{\cdot}5$

x	$f(x)=\dfrac{1}{1+x}$	x	$f(x)=\dfrac{1}{1+x}$
0·0	1·000000	2·0	0·333333
0·5	0·666667	2·5	0·285714
1·0	0·500000	3·0	0·250000
1·5	0·400000		

By Weddle's rule,

$$I = \int_0^3 \frac{dx}{1+x} = \frac{3}{10}(0{\cdot}5)\ \{1{\cdot}000000 + 5\,(0{\cdot}666667) + 0{\cdot}500000$$
$$+ 6\,(0{\cdot}400000) + (0{\cdot}333333) + 5\,(0{\cdot}285714) + (0{\cdot}250000)\} = 1{\cdot}3867857$$

Also $\quad I = \displaystyle\int_0^3 \frac{dx}{1+x} = |\log(1+x)|_0^3 = \log_e 4 = 2\log_e 2$

$$\therefore \quad 2\log_e 2 = 1{\cdot}3867857$$
$$\therefore \quad \log_e 2 = 0{\cdot}69339285 \simeq 0{\cdot}693.$$

Ex. 5-17. *Applying 'Simpson's one-third rule' evaluate*

$$\int_0^1 \frac{dx}{1+x}$$

correct to three decimal places.

Sol. Divide the range of integration into 10 equal parts by taking $h = 0{\cdot}1$

x	$f(x)=\dfrac{1}{1+x}$	x	$f(x)=\dfrac{1}{1+x}$
0·0	1·000000	0·6	0·625000
0·1	0·909091	0·7	0·588235
0·2	0·833333	0·8	0·555556
0·3	0·769231	0·9	0·526316
0·4	0·714286	1·0	0·500000
0·5	0·666667		

$$\therefore \quad I = \int_0^1 \frac{dx}{1+x} = \frac{h}{3}\{(y_0+y_{10}) + 4(y_1+y_3+y_5+y_7+y_9) + 2\,(y_2+y_4+y_6+y_8)\}$$
$$= \frac{0{\cdot}1}{3}\{1{\cdot}500000 + 4(3{\cdot}459540) + 2(2{\cdot}728175)\}$$
$$= 0{\cdot}69315$$
$$= 0{\cdot}693.$$

Ex. 5-18. *Evaluate* $I = \displaystyle\int_{0\cdot2}^{1\cdot4} (\sin x - \log x + e^x)\,dx$

by (1) *Simpson's rule*
(2) *Weddle's rule*

taking $h = 0{\cdot}1$ *in each case.*

Sol. Let $y = \sin x - \log x + e^x$

x	$\sin x$	$\log x$	e^x	y
0·2	0·198669	– 1·609438	1·221403	3·029510
0·3	0·295520	– 1·203973	1·349859	2·849352
0·4	0·389418	– 0·916291	1·491825	2·797534
0·5	0·479426	– 0·693147	1·648721	2·821294
0·6	0·564642	– 0·510826	1·822119	2·897587
0·7	0·644218	– 0·356675	2·013753	3·014646
0·8	0·717356	– 0·223143	2·225541	3·166040
0·9	0·783327	– 0·105360	2·459603	3·348290
1·0	0·841471	– 0·000000	2·718282	3·559753
1·1	0·891207	– 0·095310	3·004166	3·800063
1·2	0·932039	– 0·182322	3·320117	4·069834
1·3	0·963558	– 0·262364	3·669297	4·370491
1·4	0·985450	– 0·336472	4·055200	4·704178

(1) By Simpson's one-third rule,

$$I = \frac{h}{3}\,[(y_0+y_{12}) + 4(y_1+y_3+y_5+y_7+y_9+y_{11}) + 2(y_2+y_4+y_6+y_8+y_{10})]$$

$$= \frac{0\cdot 1}{3}\,[7{\cdot}733688 + 80{\cdot}816544 + 32{\cdot}981496] = 4{\cdot}0510576 \simeq 4{\cdot}05106$$

(2) Weddle's rule.

$$I = \frac{3h}{10}\,[y_0+y_2+y_4+y_8+y_{10}+y_{12}) + 2y_6 + 5(y_1+y_5+y_7+y_{11}) + 6(y_3+y_9)]$$

$$= \frac{3(0\cdot 1)}{10}\,[21{\cdot}058396 + 6{\cdot}332080 + 67{\cdot}913895 + 39{\cdot}728142]$$

$$= 4{\cdot}05097539 \simeq 4{\cdot}05098.$$

Ex. 5-19. *Show that* $\int_0^1 y_x\,dx = \frac{1}{12}\,(5y_1 + 8y_0 - y_{-1}).$ *(approximately)*

Sol. By Lagrange's formula

$$y_x \simeq \frac{(x-0)(x-1)}{(-1-0)(-1-1)}\,y_{-1} + \frac{(x+1)(x-1)}{(0+1)(0-1)}\,y_0 + \frac{(x+1)(x-0)}{(1+1)(1-0)}\,y_1$$

$$\simeq \frac{1}{2}\,(x^2-x)\,y_{-1} - (x^2-1)\,y_0 + \frac{1}{2}\,(x^2+x)\,y_1$$

$$\int_0^1 y_x\,dx \simeq \frac{1}{2}\left(\frac{1}{3}-\frac{1}{2}\right)y_{-1} - \left(\frac{1}{3}-1\right)y_0 + \frac{1}{2}\left(\frac{1}{3}+\frac{1}{2}\right)y_1$$

$$\simeq \frac{1}{12}\,(5y_1 + 8y_0 - y_{-1}).$$

Ex. 5-20. *Show that*

$$\int_{-1/2}^{1/2} y_x\,dx = \frac{1}{2}\,\{y_{-1/2}+y_{1/2}\} + \frac{1}{24}\,\{\Delta y_{-3/2} - \Delta y_{1/2}\} \quad \textit{(approximately)}$$

Sol. By Lagrange's formu

$$\int_{-1/2}^{1/2} y_x\,dx \simeq -\frac{1}{24}\,y_{-}$$

$$\simeq \frac{1}{2}\,\{y_{-1/}$$

$$\simeq \frac{1}{2}\,\{y_{-1/}$$

1. Find the first derivative

x : 1·00

$f(x)$: 0·841471 0·

2. Find the values of $f'(10$

x : 10

$f(x)$: 3·162278 3·

3. Find the value of cos 1·7

x	$\sin x$
1·70	0·99166
1·72	0·98888
1·74	0·98571
1·76	0·98215
1·78	0·97819
1·80	0·97384

e^x	y
:21403	3·029510
:49859	2·849352
:91825	2·797534
:48721	2·821294
:22119	2·897587
)13753	3·014646
:25541	3·166040
:59603	3·348290
'18282	3·559753
)04166	3·800063
:20117	4·069834
:69297	4·370491
)55200	4·704178

$+ y_{11}) + 2(y_2 + y_4 + y_6 + y_8 + y_{10})]$

$81496] = 4{\cdot}0510576 \simeq 4{\cdot}05106$

$5(y_1 + y_5 + y_7 + y_{11}) + 6(y_3 + y_9)]$

$913895 + 39{\cdot}728142]$

(approximately)

$y_0 + \dfrac{(x+1)(x-0)}{(1+1)(1-0)} y_1$

$+ x) y_1$

$+ \dfrac{1}{2}\Big) y_1$

$'_{1/2}\}$ *(approximately)*

Sol. By Lagrange's formula

$$y_x = \frac{\left(x+\frac{1}{2}\right)\left(x-\frac{1}{2}\right)\left(x-\frac{3}{2}\right)}{\left(-\frac{3}{2}+\frac{1}{2}\right)\left(-\frac{3}{2}-\frac{1}{2}\right)\left(-\frac{3}{2}-\frac{3}{2}\right)} y_{-3/2}$$

$$+ \frac{\left(x+\frac{3}{2}\right)\left(x-\frac{1}{2}\right)\left(x-\frac{3}{2}\right)}{\left(-\frac{1}{2}+\frac{3}{2}\right)\left(-\frac{1}{2}-\frac{1}{2}\right)\left(-\frac{1}{2}-\frac{3}{2}\right)} y_{-1/2} + \frac{\left(x+\frac{3}{2}\right)\left(x+\frac{1}{2}\right)\left(x-\frac{3}{2}\right)}{\left(\frac{1}{2}+\frac{3}{2}\right)\left(-\frac{1}{2}+\frac{1}{2}\right)\left(\frac{1}{2}-\frac{3}{2}\right)} y_{1/2}$$

$$+ \frac{\left(x+\frac{3}{2}\right)\left(x+\frac{1}{2}\right)\left(x-\frac{1}{2}\right)}{\left(\frac{3}{2}+\frac{3}{2}\right)\left(\frac{3}{2}+\frac{1}{2}\right)\left(\frac{3}{2}-\frac{1}{2}\right)} y_{3/2}$$

$$\simeq -\frac{1}{6}\left(x^3 - \frac{3}{2}x^2 - \frac{1}{4}x + \frac{3}{8}\right) y_{-3/2}$$

$$+ \frac{1}{2}\left(x^3 - \frac{1}{2}x^2 - \frac{9}{4}x + \frac{9}{8}\right) y_{-1/2} - \frac{1}{2}\left(x^3 + \frac{1}{2}x^2 - \frac{9}{4}x - \frac{9}{8}\right) y_{1/2}$$

$$+ \frac{1}{6}\left(x^3 + \frac{3}{2}x^2 - \frac{1}{4}x - \frac{3}{8}\right) y_{3/2}$$

$$\int_{-1/2}^{1/2} y_x\, dx \simeq -\frac{1}{24} y_{-3/2} + \frac{13}{24} y_{-1/2} + \frac{13}{24} y_{1/2} - \frac{1}{24} y_{3/2}$$

$$\simeq \frac{1}{2}\{y_{-1/2} + y_{1/2}\} + \frac{1}{24}\left\{\left(y_{-1/2} - y_{-3/2}\right) - \left(y_{3/2} - y_{1/2}\right)\right\}$$

$$\simeq \frac{1}{2}\{y_{-1/2} + y_{1/2}\} + \frac{1}{24}\{\Delta y_{-3/2} - \Delta y_{1/2}\}.$$

EXERCISES

1. Find the first derivative of the function tabulated below at the point $x = 1{\cdot}002$.

x :	1·00	1·01	1·02	1·03	1·04	1·05
$f(x)$:	0·841471	0·846832	0·852108	0·857299	0·862404	0·867423

2. Find the values of $f'(10)$, $f'(15)$ and $f'(12)$ from the following table :

x :	10	11	12	13	14	15
$f(x)$:	3·162278	3·316625	3·464102	3·605551	3·741657	3·872983

[**Ans.** 0·158109, 0·1291 : 0·144337]

3. Find the value of $\cos 1{\cdot}76 = \left[\dfrac{d}{dx}\{\sin x\}\right]_{x\,=\,1{\cdot}76}$ using the following table :

x	$\sin x$	x	$\sin x$
1·70	0·99166481	1·82	0·96910913
1·72	0·98888977	1·84	0·96398300
1·74	0·98571918	1·86	0·95847128
1·76	0·98215432	1·88	0·95257619
1·78	0·97819661	1·90	0·94630009
1·80	0·97384763		

[**Ans.** 0·18807675]

4. Find the values of $f'(0\cdot6)$ and $f''(0\cdot6)$ from the following table :

x :	0·4	0·5	0·6	0·7	0·8
$f(x)$:	1·5836494	1·7974426	2·0442376	2·3275054	2·6510818

[**Ans.** 2·644225; 3·64424]

5. Find the value of $f'(1\cdot0)$, $f''(1\cdot0)$, $f'''(1\cdot0)$, $f'(1\cdot10)$, $f''(1\cdot10)$ and $f'''(1\cdot10)$ from the following table :

x	$f(x)$	x	$f(x)$
1·00	1·00000	1·20	1·09544
1·05	1·02470	1·25	1·11803
1·10	1·04881	1·30	1·14017
1·15	1·07238		

[**Ans.** 0·50024; – 0·256; 0·4; 0·4767; – 0·216; 0·320]

6. Find the values of $f'(1\cdot72)$, $f'(1\cdot7)$, $f'(1\cdot5)$, $f'(2\cdot0)$, $f''(1\cdot7)$, $f'(1\cdot5)$, and $f''(2\cdot0)$ from the following table (For central values use Stirling's formula).

x :	1·5	1·6	1·7	1·8	1·9	2·0
$f(x)$:	0·405465	0·470004	0·530628	0·587787	0·641854	0·693147

[**Ans.** 0·581393; 0·588230; 0·666702; 0·500027; – 0·345858; – 0·445391; – 0·248809]

7. Find the value of $f'(0\cdot425)$, $f'(0\cdot65)$, $f''(0\cdot425)$ and $f''(0\cdot65)$ from the following table :

x :	0·4	0·5	0·6	0·7	0·8
$f(x)$:	0·389418	0·479426	0·564642	0·644218	0·717356

[**Ans.** 0·911056; 0·796092; – 0·412735; – 0·604942]

8. Prove (*i*) of Ex. 5-3 (*a*) by the method of Ex. 5-4 (*b*) by using operators.
9. Prove (*i*) of Ex. 5-5 (*a*) by the method of Ex. 5-4 (*b*) by using operators.
10. Find the expressions of $f''(x_0)$ and $f'''(x_0)$ in terms of backward differences.
11. Find $\frac{dy}{dx}$ at $x = 1$ from the following table constructing a central difference table :

x :	1	2	3	4	5	6
y :	198669	295520	389418	479425	564642	644217

12. Show that

$$y' = \frac{1}{h}\left\{\delta_y - \frac{1}{24}\,\delta_y^3 + \frac{6}{640}\,\delta^4 y.....\right\}$$

and

$$y'' = \frac{1}{h^2}\left(\delta_y^2 - \frac{1}{12}\,\delta_y^4 + ...\right)$$

13. Evaluate $I = \int_2^{10} \frac{dx}{1+x}$ by dividing the range into eight equal parts. [**Ans.** 1·299]
14. Evaluate $I = \int_1^5 \frac{dx}{x}$ by Simpson's rule. [**Ans.** 1·62]
15. Evaluate $I = \int_4^{5\cdot2} \frac{1}{x}\,dx$ by (1) Simpson's rule (2) Weddle's rule. (Taking $h = 0\cdot2$). [**Ans.** 0·262364; 0·262364]
16. Evaluate $\int_{30^\circ}^{90^\circ} \log_{10} \sin x\, dx$ by Simpson's rule (taking ten subintervals). [**Ans.** – 0·095]

17. Using Simpson's rule and
x : 0·5 0·6
y : 0·4804 0·5669
Evaluate the integrals
(*i*) $\int_{0\cdot5}^{1\cdot1} xy\,dx$ (*ii*) $\int_{0\cdot5}^{1\cdot1} y$
18. Compute $I = \int_4^{5\cdot2} \log_e x dx$
19. Evaluate $\int_0 e^x\,dx$ using S
x : 0
e^x : 1
20. Evaluate $\int_0^{0\cdot3} (1-8x^3)^{1/}$
21. Compute $\int_0^{\pi/2} \sin x\,dx$ by (1
22. If $U_x = a + bx + cx^2$, prov
$\int_1^3 U_x dx =$
23. Calculate by Simpson's or equidistant ordinates.
24. Find an approximate value
x : 1·00
y : 3·953
25. Use Simpson's rule to pro

17. Using Simpson's rule and the table :

x :	0·5	0·6	0·7	0·8	0·9	1·0	1·1
y :	0·4804	0·5669	0·6490	0·7262	0·7985	0·8658	0·9281

Evaluate the integrals

(*i*) $\int_{0.5}^{1.1} xy\,dx$ (*ii*) $\int_{0.5}^{1.1} y^2\,dx$ (*iii*) $\int_{0.5}^{1.1} x^2 y\,dx$ (*iv*) $\int_{0.5}^{1.1} y^3\,dx$.

[**Ans.** 0·3585, 0·3210, 0·3104, 0·2444]

18. Compute $I = \int_{4}^{5.2} \log_e x\,dx$ by (*i*) Simpson's rule (*ii*) Weddle's rule (Taking $h = 0.2$).

[**Ans.** 1·827847]

19. Evaluate $\int_{0}^{4} e^x\,dx$ using Simpson's rule and the following data:

x :	0	1	2	3	4
e^x :	1	2·72	7·39	20·09	54·60

20. Evaluate $\int_{0}^{0.3} (1-8x^3)^{1/4}\,dx$ using Simpson's three-eighth rule. [**Ans.** 0·29159]

21. Compute $\int_{0}^{\pi/2} \sin x\,dx$ by (1) Trapezoidal rule (2) Simpson's rule. (using 11 ordinates).

[**Ans.** 0·9981, 1·0006]

22. If $U_x = a + bx + cx^2$, prove that

$$\int_{1}^{3} U_x dx = 2u_2 + \frac{1}{12}(u_0 - 2u_2 + u_4)$$

23. Calculate by Simpson's one-third rule an approximate value of $\int_{-3}^{3} x^4\,dx$ by taking 7 equidistant ordinates.

24. Find an approximate value of $\int_{1.00}^{1.04} y\,dx$ given the following values :

x :	1·00	1·01	1·02	1·03	1·04
y :	3·953	4·066	4·182	4·30(	4·421

25. Use Simpson's rule to prove that $\log_e 7$ is approximately 1·95.

□□

6

Curve Fitting and Method of Least Squares

6.1. Introduction

Fitting of curves to a set of numerical data is of considerable importance—Theoretical as well as practical. It is based on the following principle known as principle of least squares.

Principle of Least Squares

It says that the best or most probable value of the measured quantity is that value for which the sum of the squares of the errors is least.

6.2. Solving System of Linear Equations

Consider the system of equations

$$a_{11} x_1 + a_{12} x_2 + \ldots\ldots + a_{1n} x_n = b_1$$
$$a_{21} x_1 + a_{22} x_2 + \ldots\ldots a_{2n} x_n = b_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1} x_1 + a_{m2} x_2 + \ldots\ldots + a_{mn} x_n = b_m$$

In the case when $m > n$, the equations are solved by writing

$$S = \sum_{i=1}^{m} (b_i - a_{i1}x_1 - a_{i2}x_2 \ldots\ldots - a_{in} x_n)^2$$

and minimizing S. From calculus, the equations determining $x_1, x_2 \ldots\ldots x_n$ so that S is minimum are

$$\frac{\partial S}{\partial x_1} = 0, \ \frac{\partial S}{\partial x_2} = 0, \ldots\ldots, \frac{\partial S}{\partial x_n} = 0$$

These equations are called **Normal Equations** and the values of $x_1, x_2, \ldots\ldots, x_n$ obtained from these are called **best or most plausible values.**

Ex. 6-1. *Form the normal equations and hence find the most plausible values of x and y from the following.*

$$x + y = 3{\cdot}01, \quad 2x - y = 0{\cdot}03, \quad x + 3y = 7{\cdot}03, \quad 3x + y = 4{\cdot}97.$$

Sol. Let $S = (x + y - 3{\cdot}01)^2 + (2x - y - 0{\cdot}03)^2 + (x + 3y - 7{\cdot}03)^2 + (3x + y - 4{\cdot}97)^2$

Normal equations are

$$\frac{\partial S}{\partial x} = 0 = \frac{\partial S}{\partial y}$$

Now $\qquad \dfrac{\partial S}{\partial x} = 0$ imply

$$2(x + y - 3{\cdot}01) + 4(2x - y - 0{\cdot}03) + 2(x + 3y - 7{\cdot}03) + 6(3x + y - 4{\cdot}97) = 0.$$

i.e., $\qquad 15x + 5y - 25{\cdot}01 = 0 \qquad \ldots(1)$

ɹeast Squares

ɩle importance—Theoretical as principle of least squares.

ed quantity is that value for

ng

$x_2 \ldots. x_n$ so that S is minimum

ues of $x_1, x_2, \ldots., x_n$ obtained

ost plausible values of x and

$= 4{\cdot}97.$

$\cdot 7{\cdot}03)^2 + (3x + y - 4{\cdot}97)^2$

$) + 6(3x + y - 4{\cdot}97) = 0.$ ...(1)

and $\dfrac{\partial S}{\partial y} = 0$ imply

$$2(x + y - 3{\cdot}01) - 2(2x - y - 0{\cdot}03) + 6(x + 3y - 7{\cdot}03) + 2(3x + y - 4{\cdot}97) = 0.$$

$$5x + 12y - 29{\cdot}04 = 0 \qquad \ldots(2)$$

Solving (1) and (2)

$$x = 0{\cdot}9995, \quad y = 2{\cdot}0035.$$

Ex. 6-2. *Find the most plausible values of x, y and z from the equations given below :*

$$x + 2y + z = 1, \quad 2x + y + z = 4, \quad -x + y + 2z = 3 \quad \text{and} \quad 4x + 2y - 5z = -7.$$

Sol. Let $S = (x + 2y + z - 1)^2 + (2x + y + z - 4)^2 + (-x + y + 2z - 3)^2 + (4x + 2y - 5z + 7)^2$

Normal equations are

$$0 = \frac{\partial S}{\partial x} = 2(x + 2y + z - 1) + 4(2x + y + z - 4) - 2(-x + y + 2z - 3) + 8(4x + 2y - 5z + 7).$$

i.e., $22x + 11y - 19z + 22 = 0$ ...(1)

$$0 = \frac{\partial S}{\partial y} = 4(x + 2y + z - 1) + 2(2x + y + z - 4) + 2(-x + y + 2z - 3) + 4(4x + 2y - 5z + 7)$$

i.e., $11x + 10y - 5z + 5 = 0$ ...(2)

and $$0 = \frac{\partial S}{\partial z} = 2(x + 2y + z - 1) + 2(2x + y + z - 4) + 4(-x + y + 2z - 3) - 10(4x + 2y - 5z + 7) \qquad \ldots(3)$$

i.e., $-19x - 5y + 31z - 46 = 0$ ...(3)

From (1), (2) and (3)

$$x = 1{\cdot}17, \quad y = -0{\cdot}75, \quad z = 2{\cdot}08.$$

Ex. 6-3. *Find the most plausible values of x, y and z from the following equations :*

$$x - y + 2z = 3, \quad 3x + 2y - 5z = 5, \quad 4x + y + 4z = 21, \quad -x + 3y + 3z = 14.$$

Sol. Let $S = (x - y + 2z - 3)^2 + (3x + 2y - 5z - 5)^2 + (4x + y + 4z - 21)^2 + (-x + 3y + 3z - 14)^2$

Normal equations are

$$0 = \frac{\partial S}{\partial x} = 2(x - y + 2z - 3) + 6(3x + 2y - 5z - 5) + 8(4x + y + 4z - 21) - 2(-x + 3y + 3z - 14)$$

i.e., $27x + 6y - 88 = 0$ ...(1)

$$0 = \frac{\partial S}{\partial y} = -2(x - y + 2z - 3) + 4(3x + 2y - 5z - 5) + 2(4x + y + 4z - 21) + 6(-x + 3y + 3z - 14)$$

i.e., $6x + 15y + z - 70 = 0$ ...(2)

$$0 = \frac{\partial S}{\partial z} = 0 = 4(x - y + 2z - 3) - 10(3x + 2y - 5z - 5) + 8(4x + y + 4z - 21) + 6(-x + 3y + 3z - 14)$$

i.e., $y + 54z - 107 = 0$ ...(3)

From (1), (2) and (3)

$$x = 2{\cdot}47, \quad y = 3{\cdot}55, \quad z = 1{\cdot}92.$$

Ex. 6-4. *A man is three times as old as his son. Ten years hence his age will be twice the age of his son. Five years before the age of the man was five times that of his son. Find their present ages.*

Sol. Let x and y be the ages of the man and his son.

Then
$$x = 3y,$$
$$x + 10 = 2(y + 10) \qquad i.e., \quad x - 2y - 10 = 0$$
and
$$x - 5 = 5(y - 5) \qquad i.e., \quad x - 5y + 20 = 0$$
Let
$$S = (x - 3y)^2 + (x - 2y - 10) + (x - 5y + 20)^2$$
Normal equations are
$$0 = \frac{\partial S}{\partial x} = 2(x - 3y) + 2(x - 2y - 10) + 2(x - 5y + 20)$$
i.e.,
$$3x - 10y + 10 = 0 \qquad ...(1)$$
and
$$0 = \frac{\partial S}{\partial y} = -6(x - 3y) - 4(x - 2y - 10) - 10\,(x - 5y + 20)$$
i.e.,
$$-10x + 38y - 80 = 0 \qquad ...(2)$$
From (1) and (2),
$$x = 30, \quad y = 10.$$

6.3. The Method of Least Squares

Let $y = f(x)$, be the formula (containing m unknown parameters $a_1, a_2 \,....\, a_m$), whose form is to be inferred from the results of experiment or observation and in which the unknown parameters are to be determined from experimental or observational data $(x_1, y_1), (x_2, y_2) \,...\, (x_n, y_n)$ $(n > m)$. These sets of simultaneous values of x and y would, when substituted in the formula, give n equations in m unknowns $a_1, a_2, \,...\, a_m$ and from these equations the best values of $a_1, a_2 .... a_m$ are to be obtained. To solve this problem method of Least Squares is used.

The method of Least Squares says that the best representative formula is that for which the sum of the squares of the residuals (*i.e.*, most probable value – measured value) is minimum. Since the squares of the residuals are positive, the requirement that their sum shall be as small as possible ensures that the numerical values of the residuals will be small.

Let $\gamma_1 = f(x_i)$

$\therefore$ Residual for $x = x_i = \gamma_i - y_i$

$\therefore$ By the method of Least Squares $a_1, a_2, \,....\, n_m$ are to be obtained so that
$$S = \sum_{i=1}^{n} \{\gamma_i - y_i\}^2 = \sum_{i=1}^{n} \{f(x_i) - y_i\}^2$$
is minimum. From calculus the equations determining $a_1, a_2 \,....\, a_m$ are
$$\frac{\partial S}{\partial a_1} = 0 = \frac{\partial S}{\partial a_2} = \,.....\, = \frac{\partial S}{\partial a_m}$$
These equations are called **Normal Equations.**

6.4. Curve Fitting

(1) Fitting of Parabolic Curves.

To derive the least square equations for fitting a curve of the type.
$$y = a_0 + a_1 x + a_2 x^2 + \,.....\, + a_m x^m \qquad (a_m \neq 0)$$
to a set of n points.

Let $(x_i, y_i)\ i = 1, 2, \ldots\ldots n$ be the given data and

$$\gamma_i = a_0 + a_1 x_i + a_2 x_i^2 + \ldots\ldots + a_m x_i^m$$

Let
$$S = \sum_{i=1}^{n} (\gamma_i - y_i)^2 = \sum_{i=1}^{n} \{a_0 + a_1 x_i + \ldots. + a_m x_i^m - y_i\}^2$$

By the method of least squares S is to be minimized. Normal equations are

$$\frac{\partial S}{\partial a_0} = 0 = \frac{\partial S}{\partial a_1} = \ldots\ldots \frac{\partial S}{\partial a_m}$$

Now
$$\frac{\partial S}{\partial a_0} = \sum_{i=1}^{n} 2\{a_0 + a_1 x_i + a_2 x_i^2 + \ldots. + a_m x_i^m - y_i\}$$

$\therefore$ First normal equation is

$$\sum_{i=1}^{n} y_i = n a_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 + \ldots\ldots a_m \sum_{i=1}^{n} x_i^m \qquad \ldots(1)$$

Similarly $\dfrac{\partial S}{\partial a_1} = 0 \ldots. \dfrac{\partial S}{\partial a_m} = 0$ umply

$$\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + \ldots. + a_m \sum_{i=1}^{n} x_i^{m+1} \qquad \ldots(2)$$

$$\sum_{i=1}^{n} x_i^2 y_i = a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + \ldots. + a_m \sum_{i=1}^{n} x_i^{m+2} \qquad \ldots(3)$$

....

$$\sum_{i=1}^{n} x_i^m y_i = a_0 \sum_{i=1}^{n} x_i^m + a_1 \sum_{i=1}^{n} x_i^{m+1} + \ldots. + a_m \sum_{i=1}^{n} x_i^{2m} \qquad \ldots(m+1)$$

Eqs. (1) to $(m + 1)$ are required equations.

6.4-1. Corollary

Fitting of a straight line.

The equation of a straight line is

$$y = a_0 + a_1 x$$

$\therefore$ The normal equations are

$$\Sigma y = n a_0 + a_1 \Sigma x$$

and
$$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2$$

The calculations are simplified by taking the variable v and u (to be chosen suitably) instead of y and x. Generally, u is chosen s.t. $\Sigma u = 0$.

Ex. 6-4-1.1. Show that the best-fitting linear function for the points (x_1, y_1); (x_2, y_2);

(x_n, y_n) may be expressed in the form

$$\begin{vmatrix} x & y & 1 \\ \Sigma x_i & \Sigma y_i & n \\ \Sigma x_i^2 & \Sigma x_i y_i & \Sigma x_i \end{vmatrix} = 0.$$

Ex. 6-4-1.2. Using the principle of least squares, fit a straight line to the pairs of values $(x_i, y_i)\ i = 1, 2, \ldots. n$ with a view to determine y for given x.

Simplify the equations when the possible values of x are

$$0, \pm 1, \pm 2, \ldots\ldots. \pm k \quad \text{and} \quad n = 2k + 1$$

Ex. 6-5. *Fit a straight line trend by the method of least squares to the following data :*

Year	*Milk consumption (Million litres)*
1990	102·3
1991	101·9
1992	105·8
1993	112·0
1994	114·8
1995	118·7
1996	124·5
1997	102·9

Sol. Let x and y be the variables for years and Milk consumption.

x	u	y	v	u^2	uv	
1990	– 4	102·3	– 9·7	16	38·8	
1991	– 3	101·9	– 10·1	9	30·3	
1992	– 2	105·8	– 6·2	4	12·4	
1993	– 1	112·0	0	1	0	$u = x - 1994$
1994	0	114·8	2·8	0	0	$v = y - 112$
1995	1	118·7	6·7	1	6·7	
1996	2	124·5	12·5	4	25·0	
1997	3	102·9	– 9·1	9	– 27·3	
	– 4		– 13·1	44	85·9	

Let the straight line to be fitted be

$$v = a + bu$$

where the co-efficients 'a' and 'b' are to be determined from the normal equations.

$$\Sigma v = na + b\Sigma u$$

$$\Sigma uv = a\Sigma u + b\Sigma u^2$$

Substituting the values of Σv, Σu etc.

$$-13{\cdot}1 = 8a - 4b$$

$$85{\cdot}9 = (-4a) + 44b$$

$$\therefore \qquad b = 1{\cdot}89, \; a = -0{\cdot}69$$

$\therefore$ Eq. of the straight line is

$$v = -0{\cdot}69 + 1{\cdot}89\,u$$

i.e., $\qquad y - 112 = -0{\cdot}69 + 1{\cdot}89\,(x - 1994).$

Ex. 6-6. *Fit a straight line trend by the method of least squares to the following series.*

Year	*Price Index*
1991	107
1992	110
1993	114
1994	112
1995	115
1996	113

Sol. Let x and y be the variables for years and Price Index.

Years x	u	*Price Index*	v	u^2	uv	
1991	-3	107	-5	9	15	
1992	-2	110	-2	4	4	
1993	-1	114	2	1	-2	$u = (x - 1994)$
1994	0	112	0	0	0	$v = (y - 112)$
1995	1	115	3	1	3	
1996	2	113	1	4	2	
	-3		-1	19	22	

Let the equation of the straight line to be fitted be

$$v = a + bu$$

where the co-efficients 'a' and 'b' are to be determined from normal equations

$$\Sigma v = na + b\Sigma u$$

$$\Sigma uv = a\Sigma u + b\Sigma u^2$$

Substituting the values of Σu, Σv etc.

$$-1 = 6a - 3b \quad \text{...(1)}$$

and

$$22 = (-3)a + 19b \quad \text{...(2)}$$

Multiplying (2) by '2' and adding to (1) we get

$$43 = 35\,(b)$$

or

$$b = 1{\cdot}23$$

$\therefore$ From (1)

$$6a = 2{\cdot}69$$

or

$$a = 0{\cdot}45$$

$\therefore$ The equation is

$$v = 0{\cdot}45 + 1{\cdot}23u$$

or

$$y - 112 = 0{\cdot}45 + 1{\cdot}23\,(x - 1994).$$

Ex. 6-7. *Compute the straight line trend equation for the data below by the method of least squares; determine the annual trend estimates of each year.*

Year	*Sales in thousands of Rs.*
1999	25
2000	30
2001	40
2002	50
2003	45

Sol. Let x and y be the variables for years and sales.

Years x	*Sales* y	u	v	u^2	uv	Total *Values*	
1999	25	-2	-15	4	30	26	
2000	30	-1	-10	1	10	32	$u = (x - 2001)$
2001	40	0	0	0	0	38	
2002	50	1	10	1	10	44	$v = (y - 40)$
2003	45	2	5	4	10	50	
		0	-10	10	60		

Let the equation of the straight line be

$$v = a + bu$$

where the co-efficient 'a' and 'b' are to be determined from the normal equations

$$\Sigma v = na + b\Sigma u$$
$$\Sigma uv = a\Sigma u + b\Sigma u^2$$

Substituting the values of Σu etc.

$$-10 = 5a \quad \text{or} \quad a = -2$$

and $$60 = 10b \quad \text{or} \quad b = 6$$

$\therefore$ The equation is

$$v = -2 + 6u$$
$$(y - 40) = -2 + 6(x - 2001)$$

Annual Trend estimates for each year are shown in the table.

Ex. 6-8. *Show that the line of fit to the following data is given by* $y = 0{\cdot}7x + 11{\cdot}28$:

x :	0	5	10	15	20	25
y :	12	15	17	22	24	30

Sol.

x	y	u	v	u^2	uv	
0	12	−3	−5	9	15	
5	15	−2	−2	4	4	
10	17	−1	0	1	0	$u = \frac{(x-15)}{5}$
15	22	0	5	0	0	$v = (y - 17)$
20	24	1	7	1	7	
25	30	2	13	4	26	
		−3	18	19	52	

Let the equation of the line be

$$v = a + bu$$

The normal equations are

$$\Sigma v = na + b\Sigma u$$

and $$\Sigma uv = a\Sigma u + b\Sigma u^2$$

Substituting the values of Σu etc.

$$18 = 6a - 3b \qquad ...(1)$$
$$52 = -3a + 19b \qquad ...(2)$$

$\therefore$ $$b = 3{\cdot}486$$
$$a = 4{\cdot}743$$

$\therefore$ The equation to the straight line is

$$v = 4{\cdot}743 + 3{\cdot}486u$$

or $$y - 17 = 4{\cdot}743 + 3{\cdot}486\left(\frac{x-15}{5}\right)$$

or $$y = 0{\cdot}7x + 11{\cdot}28.$$

6.4-2. Corollary

Fitting of a second degree parabola.

The equation of a second degree parabola is

$$y = a_0 + a_1 x + a_2 x^2$$

$\therefore$ The normal equations are

$$\Sigma y = na_0 + a_1\Sigma x + a_2\Sigma x^2$$
$$\Sigma xy = a_0\Sigma x + a_1\Sigma x^2 + a_2\Sigma x^3$$

and $\Sigma x^2 y =$ (

Ex. 6-9. *Fit a parabolic* ... *independent variable* :

x : 0

y : 1

Find out the difference b ... *the fitted curve when* $x = 2$.

Sol.

x	y	u	v
0	1	−2	−1·5
1	1·8	−1	−0·7
2	1·3	0	−1·2
3	2·5	1	0
4	6·3	2	3·8
		0	0·4

Let the equation of the p

The normal equations ar

Substituting the values o

or

$\therefore$ The equation of the s

$\therefore$ Value of y for $x = 2$

Also, actual value of y f

$\therefore$ Difference $= 1{\cdot}3 -$

Ex. 6-10. *Fit a paraboli*

x : 1·0

y : 1·1

Sol.

x	u	y	v
1·0	−3	1·1	−16
1·5	−2	1·3	−14
2·0	−1	1·6	−11
2·5	0	2·0	−7
3·0	1	2·7	0
3·5	2	3·4	7
4·0	3	4·1	14
	0		−27

and $$\Sigma x^2 y = a_0\Sigma x^2 + a_1\Sigma x^3 + a_2\Sigma x^4.$$

Ex. 6-9. *Fit a parabolic curve of second degree to the following data taking x as the independent variable :*

x :	0	1	2	3	4
y :	1	1·8	1·3	2·5	6·3

Find out the difference between the actual value of y and the value of y obtained from the fitted curve when x = 2.

Sol.

x	y	u	v	u^2	u^3	u^4	uv	vu^2	
0	1	– 2	– 1·5	4	– 8	16	3·0	– 6·0	
1	1·8	– 1	– 0·7	1	– 1	1	0·7	– 0·7	$u = x - 2$
2	1·3	0	– 1·2	0	0	0	0	0	
3	2·5	1	0	1	1	1	0	0	$v = y - 2{\cdot}5$
4	6·3	2	3·8	4	8	16	7·6	15·2	
		0	0·4	10	0	34	11·3	8·5	

Let the equation of the parabola be
$$v = a + bu + cu^2$$
The normal equations are
$$\Sigma v = na + b\Sigma u + c\Sigma u^2$$
$$\Sigma vu = a\Sigma u + b\Sigma u^2 + c\Sigma u^3$$
$$\Sigma vu^2 = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4$$
Substituting the values of Σu, Σv etc.
$$0{\cdot}4 = 5a + 10c \qquad ...(1)$$
$$11{\cdot}3 = 10b$$
or $$b = 1{\cdot}13 \qquad ...(2)$$
$$8{\cdot}5 = 10a + 34c$$
$$\therefore \quad c = 0{\cdot}55, \; a = -1{\cdot}02$$
$\therefore$ The equation of the second degree parabola fitted to the given data is
$$v = -1{\cdot}02 + 1{\cdot}13u + 0{\cdot}55u^2$$
$$y = 1{\cdot}48 + 1{\cdot}13\,(x-2) + 0{\cdot}55(x-2)^2$$
$\therefore$ Value of y for $x = 2$ obtained from the fitted curve
$$= 1{\cdot}48$$
Also, actual value of y for $x = 2$ is 1·3.

$\therefore$ Difference $= 1{\cdot}3 - 1{\cdot}48 = -0{\cdot}18$.

Ex. 6-10. *Fit a parabolic curve of Regression of y on x to pairs of values :*

x :	1·0	1·5	2·0	2·5	3·0	3·5	4·0
y :	1·1	1·3	1·6	2·0	2·7	3·4	4·1

Sol.

x	u	y	v	u^2	u^3	u^4	uv	vu^2	
1·0	– 3	1·1	–16	9	– 27	81	48	– 144	
1·5	– 2	1·3	– 14	4	– 8	16	28	– 56	
2·0	– 1	1·6	– 11	1	–1	1	11	–11	$u = \dfrac{x-2{\cdot}5}{0{\cdot}5}$
2·5	0	2·0	– 7	0	0	0	0	0	
3·0	1	2·7	0	1	1	1	0	0	$v = \dfrac{y-2{\cdot}7}{0{\cdot}1}$
3·5	2	3·4	7	4	8	16	14	28	
4·0	3	4·1	14	9	27	81	42	126	
	0		– 27	28	0	196	143	– 57	

Let the curve to be fitted be

$$v = a + bu + cu^2$$

The normal equations are :

$$\Sigma v = na + b\Sigma u + c\Sigma u^2$$
$$\Sigma vu = a\Sigma u + b\Sigma u^2 + c\Sigma u^3$$
$$\Sigma u^2 v = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4$$

Substituting the values of Σu, Σv etc.

$$-27 = 7a + 28c \qquad ...(1)$$
$$143 = 28b$$

or $$b = \frac{143}{28} = 5{\cdot}107$$

$$-57 = 28a + 196c \qquad ...(2)$$

Multiplying (1) by '4' and subtracting from (2)

$$84c = 51$$

or $$c = 0{\cdot}607$$

Multiplying (1) by '7' and subtracting from (2)

$$21a = -132$$

or $$a = -6{\cdot}286.$$

∴ The equation of the second degree parabola fitted to the given data is

$$v = (-6{\cdot}286) + (5{\cdot}107)u + (0{\cdot}607)u^2$$

or $$\left(\frac{y-2{\cdot}7}{0{\cdot}1}\right) = (-6{\cdot}286) + (5{\cdot}107)\left(\frac{x-2{\cdot}5}{0{\cdot}5}\right) + (0{\cdot}607)\left(\frac{x-2{\cdot}5}{0{\cdot}5}\right)^2$$

$$y = 1{\cdot}0354 - 0{\cdot}1926x + (0{\cdot}2428)x^2.$$

Ex. 6-11. *Fit a second degree parabola to the following data, taking x as the independent variable :*

x :	1	2	3	4	5	6	7	8	9
y :	2	6	7	8	10	11	11	10	9

Sol.

x	u	y	v	u^2	u^3	u^4	uv	vu^2	
1	−4	2	−6	16	−64	256	24	−96	
2	−3	6	−2	9	−27	81	6	−18	
3	−2	7	−1	4	−8	16	2	−4	
4	−1	8	0	1	−1	1	0	0	$u = x - 5$
5	0	10	2	0	0	0	0	0	
6	1	11	3	1	1	1	3	3	
7	2	11	3	4	8	16	6	12	$v = y - 8$
8	3	10	2	9	27	81	6	18	
9	4	9	1	16	64	256	4	16	
	0		2	60	0	708	51	−69	

Let the parabola of second degree to be fitted be

$$v = a + bu + cu^2$$

The normal equations for this are :

$$\Sigma v = na + b\Sigma u + c\Sigma u^2$$
$$\Sigma vu = a\Sigma u + b\Sigma u^2 + c\Sigma u^3$$

$\Sigma vu^2 =$

Substituting the values (

$2 =$

$51 =$

or $b =$

$-69 =$

i.e., $-23 =$

Multiplying (1) by '20'

$924c =$

or $c =$

∴ From (1), $a =$

Thus the equation of the

$v =$

Substituting the values (

$(y-8) =$

$y =$

(2) Fitting of Curves (

Taking logarithm

$\log_{10} y =$

or $Y =$

(where $Y =$

The constant A and b ca

From A, a is obtained on tak

(3) Fitting of curves o

Taking logarithm

$\log_{10} y =$

or $Y =$

(where $Y =$

The constants A, B and

Ex. 6-12. *The populati*

Year : 193

Population in Millions : 3·9

By fitting a curve of the

Sol.

Year x	*Population* y
1931	3·9
1941	5·3
1951	7·3
1961	9·6
1971	12·9
1981	17·1
1991	23·2
2001	30·5

$$\Sigma vu^2 = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4$$

Substituting the values of Σu, Σv etc., in these equations :

$$2 = 9a + 60c \qquad ...(1)$$

$$51 = 60b$$

or $$b = 0{\cdot}85$$

$$-69 = 60a + 708c$$

i.e., $$-23 = 20a + 236c \qquad ...(2)$$

Multiplying (1) by '20' and (2) by '9' and subtracting

$$924c = -247$$

or $$c = -0{\cdot}267$$

$\therefore$ From (1), $$a = \left(\frac{18{\cdot}02}{9}\right) = 2{\cdot}002$$

Thus the equation of the parabola of second degree is

$$v = 2{\cdot}002 + (0{\cdot}85)u - (0{\cdot}267)u^2$$

Substituting the values of u and v

$$(y-8) = 2{\cdot}002 + (0{\cdot}85)(x-5) - (0{\cdot}267)(x-5)^2,$$

$$y = (-0{\cdot}923) + (3{\cdot}52)x - (0{\cdot}267)x^2$$

(2) Fitting of Curves of the type $y = ax^b$

Taking logarithm

$$\log_{10} y = \log_{10} a + b \log_{10} x$$

or $$\gamma = A + bX$$

(where $\gamma = \log_{10} y$, $A = \log_{10} a$, and $X = \log_{10} x$)

The constant A and b can be obtained as in (6.4-1) by using *γ and X instead of y and x. From A, a is obtained on taking antilog.*

(3) Fitting of curves of the type $y = ab^x$

Taking logarithm

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

or $$\gamma = A + xB$$

(where $\gamma = \log_{10} y$, $A = \log_{10} a$, $B = \log_{10} x$)

The constants A, B and hence a, b are obtained as in (2).

Ex. 6-12. *The population of a state at ten yearly intervals is given below :*

Year :	1931	1941	1951	1961	1971	1981	1991	2001
Population in Millions :	3·9	5·3	7·3	9·6	12·9	17·1	23·2	30·5

By fitting a curve of the from $y = ab^x$ to this data estimate the population for 2011.

Sol.

Year x	*Population* y	$\gamma = \log_{10} y$	u	u^2	$u\gamma$	
1931	3·9	0·5911	−4	16	−2·3644	
1941	5·3	0·7243	−3	9	−2·1729	
1951	7·3	0·8633	−2	4	−1·7266	
1961	9·6	0·9823	−1	1	−0·9823	$u = \dfrac{x-1971}{10}$
1971	12·9	1·1106	0	0	0	
1981	17·1	1·2330	1	1	1·2330	
1991	23·2	1·3655	2	4	2·7310	
2001	30·5	1·4843	3	9	4·4529	
		8·3544	−4	44	1·1707	

The equation to the curve to be fitted is

$$y = ab^x$$

Taking $\log_{10}$

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

or
$$\gamma = A + xB$$

where $\gamma = \log_{10} y,\quad A = \log_{10}\ a,\quad B = \log_{10} b$

Replacing x by u

$$\gamma = A + Bu$$

The normal equations are

$$\Sigma\gamma = nA + B\Sigma u$$
$$\Sigma u\gamma = A\Sigma u + B\Sigma u^2$$

Substituting the values of $\Sigma\gamma$ etc.

$$8{\cdot}3544 = 8A - 4B \quad ...(i)$$

and
$$1{\cdot}1707 = (-4)A + 44B \quad ...(ii)$$

Multiplying (*ii*) by '2' and adding to (*i*)

$$10{\cdot}6958 = 84B$$

or
$$B = 0{\cdot}12733$$

$\therefore$ From (*i*)

$$8A = 8{\cdot}86372$$
$$A = 1{\cdot}107965$$

$\therefore$ The equation is

$$\gamma = 1{\cdot}107965 + 0{\cdot}12733u$$
$$= 1{\cdot}107965 + 0{\cdot}12733\left(\frac{x-1971}{10}\right)$$

$\therefore$ For $x = 2011$,

$$\gamma = 1{\cdot}107965 + 0{\cdot}50932$$
$$= 1{\cdot}617285$$
$$= 1{\cdot}6173$$

$\therefore$
$$y = 41{\cdot}43$$

$\therefore$ Population for the year 2011 = 41·43 millions.

Ex. 6-13. (*a*) *Derive the least-square equations for fitting a curve of the type* $y = ax^2 + \dfrac{b}{x}$ *to a set of n points.*

(*b*) *Fit the curve* $y = ax^2 + \dfrac{b}{x}$ *to the data given below :*

x :	1	2	3	4
y :	$-1{\cdot}51$	0·99	3·88	7·66

Sol. (*a*) Let (x_i, y_i), $i = 1, 2, \ldots n$ be the given data and

$$Y_i = ax_i^2 + \frac{b}{x_i}$$

Let
$$S = \sum_{i=1}^{n} (Y_i - y_i)^2 = \sum_{i=1}^{n} \left\{ax_i^2 + \frac{b}{x_i} - y_i\right\}^2$$

Normal equations are

$$0 = \frac{\partial S}{\partial a} = \sum_{i=1}^{n} 2\left\{ax_i^2 + \frac{b}{x_i} - y_i\right\}x_i^2$$

i.e.,
$$\sum_{i=1}^{n} y_i x_i^2 = a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i \quad ...(1)$$

and $0 = \frac{\partial}{\partial}$

i.e., $\sum_{i=1}^{n} \frac{y_i}{x_i} = a$

(1) and (2) are required e

(*b*)

x	y	x^2
1	$-1{\cdot}51$	1
2	0·99	4
3	3·88	9
4	7·66	16
10		

Substituting values in eqs

159·9300 = 3

and 2·1933 = 1

From (3) and (4)

$a = 0$

$\therefore$ The equation of the cu

$y = 0$

Ex. 6-14. *Derive the lea*

set of n points.

Sol. Let $(x_i, y_i)i = 1, 2....$

$Y = c$

Let $S =$

Normal equations are

$\frac{\partial S}{\partial a} =$

i.e., $\sum_{i=1}^{n}$

i.e., $\sum_{i=1}^{n} x_i y_i =$

and $\frac{\partial S}{\partial b} =$

i.e., $\sum_{i=1}^{n} \frac{y_i}{x_i} =$

(1) and (2) are required

and $$0 = \frac{\partial S}{\partial b} = \sum_{i=1}^{n} 2\left\{ax_i^2 + \frac{b}{x_i} - y_i\right\}\left\{\frac{1}{x_i}\right\}$$

i.e., $$\sum_{i=1}^{n} \frac{y_i}{x_i} = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right) \qquad ...(2)$$

(1) and (2) are required equations.

(*b*)

x	y	x^2	x^4	$\frac{1}{x}$	$\frac{1}{x^2}$	yx^2	$\frac{y}{x}$
1	– 1·51	1	1	1·0000	1·0000	– 1·5100	– 1·5100
2	0·99	4	16	0·5000	0·2500	3·9600	0·4950
3	3·88	9	81	0·3333	0·1111	34·9200	1·2933
4	7·66	16	256	0·2500	0·0625	122·5600	1·9150
10			354		1·4236	159·9300	2·1933

Substituting values in eqs. (1) and (2)

$$159{\cdot}9300 = 354a + 10b \qquad ...(3)$$

and $$2{\cdot}1933 = 10a + 1{\cdot}4236b \qquad ...(4)$$

From (3) and (4)

$$a = 0{\cdot}509, \qquad b = -2{\cdot}04$$

∴ The equation of the curve best fitted to the given data is

$$y = 0{\cdot}509\,x^2 - \frac{2{\cdot}04}{x}.$$

Ex. 6-14. *Derive the least-square equations for fitting a curve of type* $y = ax + \frac{b}{x}$ *to a set of n points.*

Sol. Let $(x_i, y_i)\,i = 1, 2 .... n$ be the given data and

$$Y = ax_i + \frac{b}{x_i}$$

Let $$S = \sum_{i=1}^{n} (Y_i - y_i)^2 = \sum_{i=1}^{n}\left(ax_i + \frac{b}{x_i} - y_i\right)^2$$

Normal equations are

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{n} 2\left\{ax_i + \frac{b}{x_i} - y_i\right\}(x_i)$$

i.e., $$\sum_{i=1}^{n} (ax_i^2 + b - x_i y_i) = 0$$

i.e., $$\sum_{i=1}^{n} x_i y_i = nb + a\sum_{i=1}^{n} x_i^2 \qquad ...(1)$$

and $$\frac{\partial S}{\partial b} = 0 = \sum_{i=1}^{n} 2\left(ax_i + \frac{b}{x_i} - y_i\right)\left(\frac{1}{x_i}\right)$$

i.e., $$\sum_{i=1}^{n} \frac{y_i}{x_i} = na + b\sum_{i=1}^{n} \frac{1}{x_i^2} \qquad ...(2)$$

(1) and (2) are required equations.

Ex. 6-15. *Three independent measurements on each of three angles A, B, C of a triangle are as follows :*

A :	39·5	39·3	39·6
B :	60·3	62·2	60·1
C :	80·1	80·3	80·4

Obtain the best estimates of the three angles, taking into account the relation that sum of the angles is equal to 180°.

Sol. Let the measurements for angles A, B and C be denoted by x, y and z respectively.

Then $\Sigma x = 39{\cdot}5 + 39{\cdot}3 + 39{\cdot}6 = 118{\cdot}4$

$\Sigma y = 60{\cdot}3 + 62{\cdot}2 + 60{\cdot}1 = 182{\cdot}6$

and $\Sigma z = 80{\cdot}1 + 80{\cdot}3 + 80{\cdot}4 = 240{\cdot}8$

Let α, β and γ be the true values of angles A, B and C respectively.

Then $\gamma = 180 - \alpha - \beta$.

Let
$$S = \Sigma\{(\alpha - x)^2 + (\beta - y)^2 + (\gamma - z)^2\}$$
$$= \Sigma\,\{(\alpha - x)^2 + (\beta - y^2) + (180 - \alpha - \beta - z)^2\}$$

By the principle of least squares, S is to be minimized. Normal equations are

$$0 = \frac{\partial S}{\partial \alpha} = \Sigma[2(\alpha - x) - 2(180 - \alpha - \beta - z)]$$

i.e., $2\Sigma\alpha + \Sigma\beta = \Sigma x - \Sigma z + \Sigma 180.$

i.e., $6\alpha + 3\beta = 118{\cdot}4 - 240{\cdot}8 + 540 = 417{\cdot}6.$

i.e., $2\,\alpha + \beta = 139{\cdot}2$ ...(1)

and
$$0 = \frac{\partial S}{\partial \beta} = \Sigma\,[2(\beta - y) - 2(180 - \alpha\ - \beta - z)]$$

which implies.

$$\alpha + 2\beta = 160{\cdot}6$$

From (1) and (2)

$$\alpha = 39{\cdot}27,\ \ \beta = 60{\cdot}67$$
$$\gamma = 180 - \alpha - \beta = 80{\cdot}06. \quad ..(2)$$

EXERCISES

1. Find the most plausible values of x and y from the following equations :
$x + y = 3,\ x - y = 2,\ x + 2y - 4 = 0,\ x = 2y + 1.$ **[Ans.** $x = 2{\cdot}5, y = 0{\cdot}7$**]**

2. A and B are two brothers A is ten years older than B. Five years before A's age was twice that of B's. Five years hence twice the age of A will be same as three times that of B. Find their present ages. **[Ans. 25, 15]**

3. Fit a straight line to the data given below :

x :	6	7	7	8	8	8	9	9	10
y :	5	5	4	5	4	3	4	3	3

[Ans. $y = -\,0{\cdot}5x + 8$**]**

4. Fit a straight line to the following data treating 'y' as the dependent variable.

x :	1	2	3	4	5
y :	5	7	9	10	11

[Ans. $y = 3{\cdot}9 + 1{\cdot}5x$**]**

5. Fit a straight line to the data given below; showing the production of a commodity in different years :

Year x :	1991	1992	1993	1994	1995
Production y : (1000 tons)	10	12	8	10	14

[Ans. $y = 0{\cdot}6\,(x - 1993) + 10{\cdot}8$**]**

6. Fit a straight line to the fo
x : 1
y : 1

7. The weights of a calf tak using the method of least
Age : 1 2
Weight : 52·5 58·

8. Find the equation of the following points :
x : 0·5
y : 0·31 0

9. Below are given figures o
Year : 1991
Production : 80
(in thousand mounds)
Fit a straight line trend by

10. Fit a straight line to the fo
x : 1
y : 2·4

11. Fit a straight line to the fo
x : 1
y : 1·4

12. The profits, Rs. 100y, of a
x : 1
y : 25
Fit the second degree para
$- 3)^2$]

13. Fit a parabola of second de
x : 1
y : 1·8

14. The profits Rs. y of a cert
x : 1
y : 1250 1
Show that the parabolic re
$y = 1140 + 72x$

15. Fit a second degree parab
x : 0
y : 3·1950 3·
x : 0·6
y : 3·1807 3·

16. Fit a second degree parab
x : 0
y : 1

6. Fit a straight line to the following data regarding x as the independent variable :

x :	1	1	2	3	4
y :	1	1·8	3·3	4·5	6·3

[**Ans.** $y = 0{\cdot}72 + 1{\cdot}33x$]

7. The weights of a calf taken at weekly intervals are given below. Fit a straight line using the method of least squares and calculate the average rate of growth per week.

Age :	1	2	3	4	5	6	7	8	9	10
Weight :	52·5	58·7	65·0	70·2	75·4	81·1	87·2	95·5	102·2	108·4

[**Ans.** $y = 79{\cdot}62 + 6{\cdot}16\,(x - 5{\cdot}5)$; 6·16]

8. Find the equation of the straight line which comes nearest to passing through the following points :

x :	0·5	1·0	1·5	2·0	2·5	3·0
y :	0·31	0·82	1·29	1·85	2·51	3·02

[**Ans.** $y = -\,0{\cdot}285 + 1{\cdot}10x$]

9. Below are given figures of production of a sugar factory.

Year :	1991	1992	1993	1994	1995	1996	1997	1998
Production : (in thousand mounds)	80	90	92	83	94	99	92	100

Fit a straight line trend by the method of least squares to the above data.

[**Ans.** $y = 94 + 3(x - 1995)$]

10. Fit a straight line to the following data :

x :	1	2	3	4	5	8
y :	2·4	3	3·6	4	5	6

11. Fit a straight line to the following data :

x :	1	2	3	4	6	8
y :	1·4	3	3·6	4	5	6

12. The profits, Rs. $100y$, of a certain company in the xth year of its life are given by

x :	1	2	3	4	5
y :	25	28	33	39	46

Fit the second degree parabola of y on x. [**Ans.** $y = 32{\cdot}914 + 5{\cdot}3\,(x - 3) + 0{\cdot}643(x - 3)^2$]

13. Fit a parabola of second degree to the following data taking x as independent variable:

x :	1	2	3	4	5
y :	1·8	5·1	9·0	14	19

14. The profits Rs. y of a certain company in the xth year of its existence are given by :

x :	1	2	3	4	5
y :	1250	1400	1650	1950	2300

Show that the parabolic regression of y on x is

$$y = 1140 + 72x + 32{\cdot}15x^2$$

15. Fit a second degree parabola to the data given below :

x :	0	0·1	0·2	0·3	0·4	0·5
y :	3·1950	3·2299	3·2532	3·2611	3·2516	3·2282
x :	0·6	0·7	0·8	0·9		
y :	3·1807	3·1266	3·0594	2·9759		

[**Ans.** $y = 3{\cdot}1951 + 0{\cdot}4425x - 0{\cdot}7653x^2$]

16. Fit a second degree parabola to the following data :

x :	0	1	2	3	4
y :	1	5	0	22	38

[**Ans.** $y = 5{\cdot}914 + 9{\cdot}1\,(x - 2) + 3{\cdot}643\,(x - 2)^2$]

---

angles A, B, C of a triangle

39·6
60·1
80·4

count the relation that sum

d by x, y and z respectively.

ectively.

$\beta - z)^2\}$
rmal equations are

$- z)]$

...(1)

$- z)]$

..(2)

ving equations :
[**Ans.** $x = 2{\cdot}5, y = 0{\cdot}7$]
ve years before A's age was
l be same as three times that
[**Ans.** 25, 15]

8 9 9 10
3 4 3 3
[**Ans.** $y = -\,0{\cdot}5x + 8$]
dependent variable.

[**Ans.** $y = 3{\cdot}9 + 1{\cdot}5x$]
oduction of a commodity in

1995
14

s. $y = 0{\cdot}6\,(x - 1993) + 10{\cdot}8$]

17. Fit the curve $y = ae^{bx}$ to the data given below :

x :	0	2	4
y :	5·012	10	31·62

($e = 2{\cdot}71828$) [**Ans.** $y = 4{\cdot}642\, e^{0{\cdot}46x}$]

18. Fit the curve $y = ab^x$ to the data given below :

x :	2	3	4	5
y :	144	172·8	207·4	248·8

[**Ans.** $y = 100(1{\cdot}2)^x$]

19. Fit the curve $y = ae^{bx}$ to the following data :

x :	1	2	3	4	5	6	7	8
y :	15·3	20·5	27·4	36·6	49·1	65·6	87·8	117·6

[**Ans.** $y = 11{\cdot}58\ e^{0.2898x}$]

20. Fit $y = ae^{bx}$ to the following data :

x :	2·5	5·0	7·5	10·0	12·5	15·0
y :	76	52	35	25	16	11

[**Ans.** $y = (113{\cdot}4)\ e^{-0.1549x}$]

21. Fit the curve of the type $xy^a = b$ to the following data :

x :	0·5	1·0	1·5	2·0	2·5	3·0
y :	1·62	1·00	0·75	0·62	0·52	0·46

22. Use the method of least squares to determine 'a' and 'b' in the formula $y = ax + bx^2$ for the following data :

x :	1	2	3	4	5
y :	1·8	5·1	8·9	14·1	19·8

Calculate the value of y for $x = 2$. [**Ans.** $a = 1{\cdot}521,\ b = 0{\cdot}49,\ 5{\cdot}006$]

23. Fit $y = a + bx^3$ to the following data :

x :	5	7	9	11	12
y :	290	560	1044	1810	2300

[**Ans.** $y = 130{\cdot}71 + 1{\cdot}2572x^3$]

□□

Define Pr

7.1. Introduction

A fundamental principle in
if these experiments are repeat
results are essentially the sam
essentially the same even thoug
be called **random experiment**

Below are certain terms wh

Trial. *Performing of an ex*

Cases. *Various possible ou*

Event. *It is used to represe*

Sample space. *It is the set*

Event is a subset of sample
single members.

The class of all events ass
space.

(E.L.) Equally Likely Ca
none of them can be preferred

(M.E.) Mutually Exclus
exclusive when no two of them

Exhaustive Cases (Events
all possible outcomes of a trial.

Favourable Cases. *The c*
favourable to an event.

In a trial, there is always u
To measure this uncertainty th
number between 0 and 1. If the e

event is sure not to occur its pro

that there are 25% chances for th

Below are two definitions c

7.1.1. Mathematical Definitio

Let 'n' be the number of cas
and 'm' of these are favourable

happening of A is defined to be

[**Ans.** $y = 4\cdot642\, e^{0\cdot46x}$]

[**Ans.** $y = 100(1\cdot2)^x$]

; 6 7 8
·1 65·6 87·8 117·6
[**Ans.** $y = 11\cdot58\; e^{0.2898x}$]

:·5 15·0
6 11
[**Ans.** $y = (113\cdot4)\; e^{-0.1549x}$]

:
·5 3·0
52 0·46
'b' in the formula $y = ax + bx^2$

5
9·8
ıs. $a = 1\cdot521,\; b = 0\cdot49,\; 5\cdot006$]

12
300
[**Ans.** $y = 130\cdot71 + 1\cdot2572x^3$]

Define Probability.

7

Probability

7.1. Introduction

A fundamental principle involved in the experiments of science and engineering is that if these experiments are repeatedly performed under very nearly identical conditions the results are essentially the same. But, there are experiments in which results will not be essentially the same even though conditions may be nearly identical. Such experiments will be called **random experiments** or simply **experiments.**

Below are certain terms which will be used subsequently.

Trial. *Performing of an experiment is called trial.*

Cases. *Various possible outcomes of a trial are termed as cases.*

Event. *It is used to represent the aim with which the experiment is performed.*

Sample space. *It is the set of all possible outcomes of an experiment.*

Event is a subset of sample space and cases are its members i.e., subsets consisting of single members.

The class of all events associated with a given experiment is defined to be the event space.

(E.L.) Equally Likely Cases (Events). *Cases (Events) are called equally likely when none of them can be preferred rather than the other.*

(M.E.) Mutually Exclusive Cases (Events). *Cases (Events) are called mutually exclusive when no two of them can occur simultaneously.*

Exhaustive Cases (Events). *A set of cases (events) is said to be exhaustive if it includes all possible outcomes of a trial.*

Favourable Cases. *The cases which entail the happening of an event are said to be favourable to an event.*

In a trial, there is always uncertainty as to whether a particular event will occur or not. To measure this uncertainty the idea of probability (or chance) was introduced. It is the number between 0 and 1. If the event is sure to occur its probability is taken to be 1 and if the event is sure not to occur its probability is taken to be zero. If the probability is $\frac{1}{4}$, it means that there are 25% chances for the event to occur and 75% chances for the event not to occur.

Below are two definitions of probability of an event.

7.1.1. Mathematical Definition

Let 'n' be the number of cases which are equally likely, mutually exclusive and exhaustive and 'm' of these are favourable to the happening of an event 'A'. Then the probability of the happening of A is defined to be $\frac{m}{n}$.

7.1.2. Statistical Definition

If a trial is repeated a number of times under essentially the same conditions, then limiting value of the ratio of the number of times the event happens to the number of trials, as the number of trials increases indefinitely, is called the probability of the happening of that event. (It is assumed that the ratio approaches a finite and unique limit).

Both these definitions have serious difficulties, the first because the word "equally likely" is vague and second because of the vagueness of infinite number of trials. Because of these difficulties, the following axiomatic approach to probability was introduced.

7.1.3. Axioms of Probability

Let Ω be a sample space. Let $\mathscr{A}$ be the class of events in Ω. To each A in $\mathscr{A}$ is associated a real number $P(A)$ st.

(1) $P(A) \geq 0, \forall A \in \mathscr{A}$

(2) $P(S) = 1$

(3) If $A_1, A_2, \ldots\ldots$ are any number of mutually exclusive events in $\mathscr{A}$ then

$$P(A_1 + A_2 + \ldots) = P(A_1) + P(A_2) + \ldots$$

Obviously, P is a real valued function defined on $\mathscr{A}$.P. is called a probability function and $P(A)$ the probability of the event A and $0 \leq P(A) \leq 1$.

The triplet $(\Omega, \mathscr{A}, P[\cdot])$ is called a probability space.

Remarks. Starting with definition (7.1.1) axion (3) of (7.1.3) can be proved. In view of this $P(A)$ is taken to be $\frac{O(A)}{O(\Omega)}$, where $O(A)$ denotes the order of A.

Odds in favour of and against an event are defined as below:

$$\textit{Odds in favour of an event} = \frac{\textit{Prob. of happening}}{\textit{Prob. of non-happening}}$$

$$\textit{Odds against an event} = \frac{\textit{Prob. of non-happening}}{\textit{Prob. of happening}}$$

To find the no. of ways of getting a certain sum in rolling dice.

No. of ways of getting a sum 'r' in rolling 'n' f-faced dice

$$= \textit{co-efficient of } x^r \textit{ in } (x + x^2 + \ldots x^f)^n$$

Ex. 7-1. *The chance of an event happening is the square of the chance of a second event but the odds against the first are the cube of the odds against the second. Find the chance of each.*

Sol. Let p and p' be the chances of happening of two events.

Then $$p = p'^2$$

$$\text{Odds against the first event} = \frac{1-p}{p}$$

$$\text{and odds against the second event} = \frac{1-p'}{p'}$$

$$\therefore \quad \frac{1-p}{p} = \left(\frac{1-p'}{p'}\right)^3$$

$\therefore$

which implies $p' = \frac{1}{3}$

$\therefore$

Ex. 7-2. *In an experimen… twice as likely as w_j $(j = 1, 2$…*

Sol. Let

Then

Now,

Since, $w_1, \ldots w_n$ are exha…

$p_1 + p_2$…

$p_1(1 + 2 + \ldots$…

$\therefore$

Ex. 7-3. *The sum of two… product of two quantities is n…*

Sol. Let y be one quantit…

Then $(2n - x)$ is the oth…

Let

$$\therefore \qquad \frac{1-p'^2}{p'^2} = \frac{(1-p')^3}{p'^3}$$

which implies $p' = \dfrac{1}{3}$

$$\therefore \qquad p = \frac{1}{9}.$$

Ex. 7-2. *In an experiment there are n outcomes* $w_1, w_2, \ldots w_n$. *The outcome* w_{j+1} *is twice as likely as* w_j $(j = 1, 2 \ldots n-1)$. *Find* $P(A_k)$ *where* $A_k = (w_1, w_2, \ldots w_k)$.

Sol. Let $\qquad P(w_j) = p_j \qquad j = 1, 2 \ldots n-1$

Then $\qquad p_{j+1} = 2p_j$

Now, $\qquad p_2 = 2p_1$

$$p_3 = 2p_2 = 2^2 p_1$$

$$p_4 = 2p_3 = 2^3 p_1$$

$$p_n = 2^{n-1} p_1$$

Since, $w_1, \ldots w_n$ are exhaustive,

$$p_1 + p_2 + \ldots p_n = 1$$

$$p_1\left(1 + 2 + \ldots + 2^{n-1}\right) = 1$$

$$p_1\left(\frac{1-2^n}{1-2}\right) = 1 \Rightarrow p_1 = \frac{1}{2^n - 1}$$

$$\therefore \qquad P(A_k) = \sum_{i=1}^{k} P(w_k) = p_1 + p_2 + \ldots + p_k$$

$$= p_1\left(1 + 2 + \ldots + 2^{k-1}\right)$$

$$= p_1\left(\frac{1-2^k}{1-2}\right)$$

$$= \frac{2^k - 1}{2^n - 1}.$$

Ex. 7-3. *The sum of two positive quantities is equal to 2n. Find the chance that the product of two quantities is not less than 3/4 times of their greatest product.*

Sol. Let y be one quantity.

Then $(2n - x)$ is the other quantity

Let $\qquad y = x(2n - x)$

$$\frac{dy}{dx} = 2n - 2x$$

y is maximum when $\quad \dfrac{dy}{dx} = 0$

i.e., $\quad 2n - 2x = 0$

i.e., $\quad x = n$

$\therefore$ Maximum value of $y = n^2$

$\therefore$ x is to be such that

$$x(2n-x) \nless \frac{3}{4}n^2$$

i.e., $\quad 3n^2 - 8nx + 4x^2 \ngtr 0$

i.e., $\quad (3n-2x)(n-2x) \ngtr 0$

$\therefore \quad \dfrac{n}{2} < x < \dfrac{3n}{2}$

$\therefore$ No. of favourable cases $= \dfrac{3n}{2} - \dfrac{n}{2} = n$

and total no. of cases $= 2n$

Reqd. prob. $= \dfrac{n}{2n} = \dfrac{1}{2}$.

Ex. 7-4. *What is the chance that (a) a leap year selected at random will contain 53 Sundays, (b) a non-leap year selected at random will contain 53 Sundays ?*

Sol. (*a*) In a leap year there are 366 days *i.e.*, 52 weeks and 2 days. Remaining two days can be any two days of the week. Different possibilities are :

Sunday	and Monday
Monday	and Tuesday
Tuesday	and Wednesday
Wednesday	and Thursday
Thursday	and Friday
Friday	and Saturday
Saturday	and Sunday

In order to have 53 Sundays, out of remaining two days one must be Sunday.

No. of cases favourable to the event of having one Sunday out of 2 days = 2

Total number of cases = 7

$\therefore$ Reqd. prob. $= \dfrac{2}{7}$.

(*b*) In a non-leap year there are 365 days *i.e.*, 52 weeks and 1 days. Remaining 1 day can be any day of the week.

$\therefore$ Total no. of cases = 7.

There will be 53 Sundays if the remaining one day is Sunday.

$\therefore$ No. of favourable cases = 1

$\therefore$ Reqd. prob. $= \dfrac{1}{7}$.

at random will contain 53
3 Sundays ?
days. Remaining two days

must be Sunday.
out of 2 days = 2

days. Remaining 1 day can

y.

Ex. 7-5. *Two cards are drawn at random from a well-suffled pack of 52. Show that the chance of drawing two aces is* $\frac{1}{221}$.

Sol. Total number of ways of drawing two cards out of $52 = {}^{52}c_2$. In a pack of 52, there are 4 aces out of which 2 aces can be drawn in ${}^4c_2$ ways.

$$\therefore \text{ Reqd. prob.} = \frac{{}^4c_2}{{}^{52}c_2}$$

$$= \frac{4\cdot 3}{52\cdot 51} = \frac{1}{221}.$$

Ex. 7-6. *From a pack of 52 cards, two are drawn at random. Find the chance that one is a king and the other a queen.*

Sol. Total number of cases $= {}^{52}c_2$

Since in a pack there 4 kings and 4 queens, number of favourable cases $= {}^4c_1 . {}^4c_1$.

$$\therefore \text{ Reqd. prob.} = \frac{{}^4c_1\ {}^4c_1}{{}^{52}c_2}$$

$$= \frac{4\cdot 4}{52\cdot 51} 2\cdot 1$$

$$= \frac{8}{663}.$$

Ex. 7-7. *Four cards are drawn from a well shuffled pack of cards. What is the probability that they are from four different suits ?*

Sol. Total number of cases $= {}^{52}c_4$.

Since there is 13 cards of each suit, no of ways of drawing 4 cards belonging to different suits

$$= \left({}^{13}c_1\right)\left({}^{13}c_1\right)\left({}^{13}c_1\right)\left({}^{13}c_1\right) = 13^4$$

$$\therefore \qquad \text{Reqd. Prob.} = \frac{(13)^4}{52_{c_4}} = \frac{2197}{20825}.$$

Ex. 7-8. *From a set of 17 cards numbered 1, 2,....17, one is drawn at random. What is the chance that*

(*i*) *Its number is a multiple of 3 or 7?*

(*ii*) *Its number is multiple of 3 or 5 or both?*

Sol. (*i*) Total number of cases $= {}^{17}c_1 = 17$.

The number on the card drawn will be a multiple of 3 or 7 if it is 3, 6, 7, 9, 12, 14, or 15.

$\therefore$ Number of favourable cases = 7.

$$\therefore \qquad \text{Reqd. prob.} = \frac{7}{17}.$$

(*ii*) The number on the card drawn will be a multiple of 3 or 5 or both if it is 3, 5, 6, 9, 10, 12 or 15.

$\therefore$ Number of favourable cases = 7.

$$\therefore \text{ Reqd. prob.} = \frac{7}{17}.$$

Ex. 7-9. (*a*) *If n biscuits are distributed at random among N beggars, what is the chance that a particular beggar receives r(<n) biscuits ?*

Sol. Since one biscuit can be given in N ways, number of ways of distributing n biscuits among N beggars = N^n.

If one particular beggar receives r biscuits, remaining $(n-r)$ biscuits are to be distributed among $(N-1)$ beggars and this can be done in $(N-1)^{n-r}$ ways.

r biscuits to be given to one particular beggar can be chosen in ${}^nc_r$ ways.

$\therefore$ Number of favourable cases = ${}^nc_r.(N-1)^{n-r}$

$$\therefore \text{ Reqd. prob.} = \frac{{}^nc_r(N-1)^{n-r}}{N^n}$$

(*b*) *What is the most probable number of biscuits distributed to a particular beggar?*

Sol. Let $P(r) = \dfrac{{}^nc_r(N-1)^{n-r}}{N^n}$

Most probable number of biscuits distributed to a particular beggar is that value of r which is *s.t.*

$$P(r-1) \le P(r) \ge P(r+1)$$

Consider $$P(r-1) \le P(r)$$

i.e., $$\frac{{}^nc_{r-1}(N-1)^{n-r+1}}{N^n} \le \frac{{}^nc_r(N-1)^{n-r}}{N^n}$$

i.e., $$\frac{n!}{(r-1)!(n-r+1)!}(N-1) \le \frac{n!}{r!(n-r)!}$$

i.e., $$\frac{N-1}{n-r+1} \le \frac{1}{r}$$

i.e., $$r(N-1) \le (n+1)-r$$

i.e., $$r \le \frac{n+1}{N}$$

Consider $$P(r) \ge P(r+1)$$

i.e., $$\frac{{}^nc_r(N-1)^{n-r}}{N^n} \ge \frac{{}^nc_{r+1}(N-1)^{n-r-1}}{N^n}$$

i.e., $$\frac{n!}{r!(n-r)!}(N-1) \ge \frac{n!}{(r+1)!(n-r-1)!}$$

i.e., $$\frac{(N-1)}{n-r} \ge \frac{1}{r+1}$$

i.e.,

i.e.,

$\therefore$ Most probable value

Since r is the integer, it

In case $\dfrac{n+1}{N}$ is an integer,

Ex. 7-10. *Two different*
the probability, that the sum
their sum will exceed 13 eac
5 is $\dfrac{3}{28}$.

Sol. Total number of wa
(*i*) Different possibilitie

Number of favourable c
$\therefore$ Prob. of choosing 2 c
(*ii*) Different possibilitie

Number of favourable c
$\therefore$ Probability of choosi

(*iii*) There are only thre
chosen in ${}^3c_2$ ways.
$\therefore$ Number of favourable

$\therefore$ Reqd. prob. $= \dfrac{3}{28}$.

Ex. 7-11. *If four squares*
they should be in a diagonal
Sol. A chess board is a s
Diagonal BD divides the

i.e., $$r(N-1)+(N-1) \geq n-r$$

i.e., $$r \geq \frac{n+1}{N}-1$$

$\therefore$ Most probable value of r is such that

$$\frac{n+1}{N}-1 \leq r \leq \frac{n+1}{N}$$

Since r is the integer, it is the greatest integer less than $\frac{n+1}{N}$ if $\frac{n+1}{N}$ is not an integer.

In case $\frac{n+1}{N}$ is an integer, r can take both values $\frac{n+1}{N}$ and $\frac{n+1}{N}-1$.

Ex. 7-10. *Two different digits are chosen at random from the set 1, 2, 3,.....8. Show that the probability, that the sum of the digits will be equal to 5 is the same as the probability that their sum will exceed 13 each being* $\frac{1}{14}$. *Also show that the chance of both digits exceeding 5 is* $\frac{3}{28}$.

Sol. Total number of ways of choosing 2 digits $= {}^8c_2 = 28$.

(*i*) Different possibilities of getting 2 digits with sum 5 are :

1st-digit	2nd-digit
1	4
2	3

Number of favourable cases = 2

$\therefore$ Prob. of choosing 2 digits with sum $5 = \frac{1}{14}$.

(*ii*) Different possibilities of getting 2 digits with sum exceeding 13 are :

1st-digit	2nd-digit
6	8
7	8

Number of favourable cases = 2

$\therefore$ Probability of choosing 2 digits with sum exceeding 13

$$= \frac{1}{14}.$$

(*iii*) There are only three digits exceeding 5 and out of these three digits two can be chosen in ${}^3c_2$ ways.

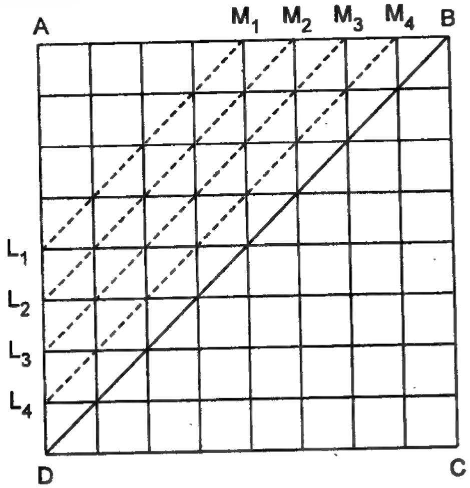
$\therefore$ Number of favourable cases $= {}^3c_2 = 3$.

$\therefore$ Reqd. prob. $= \frac{3}{28}$.

Ex. 7-11. *If four squares are chosen at random on a chessboard, find the chance that they should be in a diagonal line.*

Sol. A chess board is a square divided into 64 equal squares parallel to the sides. Diagonal BD divides the board in two equal Δs ABD and CBD. In ΔABD, four squares

along a diagonal line can be chosen along L_1M_1, L_2M_2, L_3M_3, L_4M_4 or DB which contain respectively 4, 5, 6, 7, 8 squares.



$\therefore$ Number of ways of selecting 4 squares in $\Delta\, ABD = {}^4c_4 + {}^5c_4 + {}^6c_4 + {}^7c_4 + {}^8c_4$

This is also the number of ways of selecting 4 squares in each of Δs BCD, ACD, and ABC.

$\therefore$ Total number of ways of selecting 4 squares along a diagonal line in the square

$$ABCD = 4\left\{{}^4c_4 + {}^5c_4 + {}^6c_4 + {}^7c_4\right\} + 2.\,{}^8c_4$$

$$= 364$$

(This is because the diagonals BD and AC are common to Δs ABD, BCD and ACD, ABC).

Also total number of ways of selecting 4 squares out of 64

$$= {}^{64}c_4.$$

$\therefore$ Reqd. prob. $= \dfrac{364}{64_{c_4}} = \dfrac{13}{22692}.$

Ex. 7-12. *A five-figure number is formed by the digit 0, 1, 2, 3, 4 (without repetition). Find the prob. that the number formed is divisible by 4.*

Sol. Total number of ways of arranging digits 0, 1, 2, 3, 4

$$= 5\,! = 120\,!$$

If a number start with '0' the remaining 4 digits can be arranged in 4!=24 ways and hence total number of five-figure numbers = 120 – 24 = 96.

Now the numbers ending with 04, 20, 40, 12, 24 and 32 are divisible by 4.

No. of numbers ending with 04 = Total number of ways of arranging digits 1, 2, 3 = 3! = 6.

Evidently this is also the number of numbers ending with 20 and 40 respectively.

For the numbers ending with 12, out of total number of ways of arranging remaining three digits 0, 3, 4 the number of ways in which '0' occurs first are to be discarded.

($\because$ these will give four-figure numbers).

$\therefore$ No. of numbers ending with 12 = 3! – 2! = 4.

Evidently this is also the number of numbers ending with 24 and 32 respectively.

$\therefore$ Total number of five
divisible by 4

$\therefore$ Reqd. prob.

Ex. 7-13. *Out of (2n+*
Find the chance that the nu

Sol. Total number of ca

Different possibilities c
1, 2
2, 3
3, 4
4, 5
5, 6
.....

and so on.

Number of terms in 1s
Number of terms in 2n
Number of terms in 3r
Number of terms in 4t
Number of terms in 5t

and so on.

$\therefore$ No. of favourable c

$\therefore$ Reqd. prob.

Ex. 7-14. *Four cards*
probabilities of the followi
(a) The cards are of th
etc.)
(b) There is at least or
(c) Only two of the fou

Sol. (*a*) Total number

First card (say A_1) ca

the 36 cards obtained on di

the third card (say A_3) mu

to the suits and denominat

obtained on discarding the

$\therefore$ No. of favourable c

$\therefore$ Total number of five-figure numbers, formed by the digits 0, 1, 2, 3, 4, which are divisible by 4

$$= 3(6+4) = 30$$

$$\therefore \text{ Reqd. prob.} \qquad = \frac{30}{96} = \frac{5}{16}.$$

Ex. 7-13. *Out of (2n+1) tickets consecutively numbered, three are drawn at random. Find the chance that the numbers on them are in A.P.*

Sol. Total number of cases $= {}^{2n+1}c_3$.

Different possibilities of drawing tickets with their numbers in A.P. are :

1, 2, 3; 1, 3, 5;	....	1, $(n+1)$, $(2n+1)$
2, 3, 4; 2, 4, 6;	....	2, $(n+1)$, $2n$
3, 4, 5; 3, 5, 7;	....	3, $(n+2)$, $(2n+1)$
4, 5, 6; 4, 6, 8;	....	4, $(n+2)$, $2n$
5, 6, 7; 5, 7, 9;	....	5, $(n+3)$, $(2n+1)$

....

and so on.

Number of terms in 1st sequence $= n$

Number of terms in 2nd sequence $= (n-1)$

Number of terms in 3rd sequence $= (n-1)$

Number of terms in 4th sequence $= (n-2)$

Number of terms in 5th sequence $= (n-2)$

and so on.

$$\therefore \text{ No. of favourable cases} \qquad = n + 2\{(n-1)+(n-2)+.....1\}$$

$$= n + 2\frac{(n-1)n}{2} = n^2$$

$$\therefore \text{ Reqd. prob.} \qquad = \frac{n^2}{{}^{2n+1}c_3} = \frac{3n}{4n^2-1}.$$

Ex. 7-14. *Four cards are drawn out at random from a full deck of 52. Find the probabilities of the following contingencies.*

(*a*) *The cards are of the four different suits and of different denominations (1, 2, king etc.)*

(*b*) *There is at least one ace-card.*

(*c*) *Only two of the four suits are represented.*

Sol. (*a*) Total number of *ways* for taking out 4 cards $= {}^{52}c_4$.

First card (say A_1) can be any out of 52. Then the second card (say A_2) must *be* from the 36 cards obtained on discarding the cards belonging to the suit and denomination of A_1, the third card (say A_3) must be from the 22 cards obtained on discarding the cards belonging to the suits and denominations of A_1 and A_2 and the fourth card must be from the 10 cards obtained on discarding the cards belonging to the suits and denominations of A_1, A_2 and A_3.

$$\therefore \text{ No. of favourable cases} = \left({}^{52}c_1\ {}^{36}c_1\ {}^{22}c_1 \cdot {}^{10}c_1\right) \Big/ 4!$$

($\because$ Four cards can be arranged among themselves in 4! ways).

$$\therefore \text{ Reqd. prob. } = \frac{{}^{52}c_1 \; {}^{36}c_1 \; {}^{22}c_1 \cdot {}^{10}c_1}{{}^{52}c_4 \cdot 4!} = 0{\cdot}06$$

(*b*) and (*c*) are left as exercises.

Ex. 7-15. *Out of 3n consecutive numbers 3 are selected at random. Find the chance that their sum is divisible by 3.*

Sol. Let $3n$ consecutive numbers be $p+1, \ldots p+3n$ where p is any integer. These $3n$ numbers can be arranged as follows:

$$\left.\begin{array}{lll} p+1 & p+2 & p+3 \\ p+4 & p+5 & p+6 \\ p+7 & p+8 & p+9 \\ \cdots & \cdots & \cdots \\ p+3n-2 & p+3n-1 & p+3n \end{array}\right\} \qquad \ldots(A)$$

Numbers in three columns have the property that the sum of any three numbers in any particular column is divisible by 3. Now each column consists of n numbers and hence number of ways of selecting 3 numbers from any one particular column $= {}^n c_3$.

Also the numbers in (A) are *s.t.*, if three numbers are chosen one from each column, their sum is divisible by 3.

Now number of ways of selecting three numbers from (A) one from each column $= n^3$.

$\therefore$ No. of favourable cases

$$= n^3 + 3.{}^n c_3$$

$$= \frac{n}{2}[3n^2 - 3n + 2].$$

Also total number of cases

$$= {}^{3n}c_3 = \frac{n}{2}(9n^2 - 9n + 2)$$

$$\therefore \text{ Reqd. prob. } = \frac{3n^2 - 3n + 2}{9n^2 - 9n + 2}.$$

Ex. 7-16. *If 6n tickets numbered 0, 1, 2,.......6n–1 are placed in a bag and three are drawn out, show that the chance that the sum of the numbers on them is equal to 6n is* $\dfrac{3n}{(6n-1)(6n-2)}$.

Sol. Total number of cases $= {}^{6n}c_3$.

Different possibilities of drawing tickets with the sum of their numbers equal to $6n$ are :

$0, 1, 6n-1,$	$0, 2, 6n-2;$	...........;	$0, 3n-1, 3n+1$
$1, 2, 6n-3,$	$1, 3, 6n-4;$	...........;	$1, 3n-1, 3n$
$2, 3, 6n-5,$	$2, 4, 6n-6;$	...........;	$2, 3n-2, 3n$
$3, 4, 6n-7,$	$3, 5, 6n-8;$	...........;	$3, 3n-2, (3n-1)$
$4, 5, 6n-9,$	$4, 6, 6n-10;$	...........;	$4, 3n-3, 3n-1$
....	....	....	
$2n-2, 2n-1, 2n+3;$	$2n-2, 2n, 2n+2$	;	$2n-1, 2n, 2n+1$

No. of terms in first sequen

No. of terms in 2nd sequen

No. of terms in 3rd sequen

No. of terms in 4th sequen

No. of terms in 5th sequen

and so on.

$\therefore$ Total number of ways of

$= \{(3n-1)+(3n-2)\}+\{(3$

$= \{(3n-1)+(3n-4)+(3n$

$= \frac{n}{2}\{4+3(n-1)\}+\frac{n}{2}\{2+3$

$\therefore$ Reqd. prob. $= \frac{3n^2}{6n_{c_3}} =$ — (6

Ex. 7-17. *Four different ol marked 1, 2, 3, 4. What is the corresponding to its number ?*

Sol. Total number of ways

Number of ways in which a

If three objects occupy their discussed above.

If two objects occupy their other's position *i.e.*, in only one

Since out of 4, two objects two objects can occupy their pla

If one object occupies its p ways by occupying the positions

Since out of 4, one object c one object can occupy its place.

$\therefore$ Total number of ways in

$\therefore$ Total number of ways in v to its number

$\therefore$ Reqd. prob.

Ex. 7-18. *A and B stand in persons is at random, find the ch*

Sol. There are in all 12 pers person remaining 11 persons can

$\therefore$ Total number of ways in

Out of 10 persons three are

${}^{10}c_3$ ways and can be arranged i

No. of terms in first sequence $= 3n-1$
No. of terms in 2nd sequence $= 3n-2$
No. of terms in 3rd sequence $= 3n-4$
No. of terms in 4th sequence $= 3n-5$
No. of terms in 5th sequence $= 3n-7$
and so on.

$\therefore$ Total number of ways of drawing tickets with the sum of their numbers equal to $6n$

$$= \{(3n-1)+(3n-2)\}+\{(3n-4)+(3n-5)\}+\{(3n-7)+(3n-8)\}+....+(2+1)$$

$$= \{(3n-1)+(3n-4)+(3n-7)+...+2\}+\{(3n-2)+(3n-5)+(3n-8)+....+1\}$$

$$= \frac{n}{2}\{4+3(n-1)\}+\frac{n}{2}\{2+3(n-1)\} = 3n^2$$

$$\therefore \text{ Reqd. prob. } = \frac{3n^2}{6n_{c_3}} = \frac{3n}{(6n-1)(6n-2)}.$$

Ex. 7-17. *Four different objects 1, 2, 3, 4 are distributed at random on four places marked 1, 2, 3, 4. What is the probability that none of the objects occupies the place corresponding to its number ?*

Sol. Total number of ways of distributing 4 objects on 4 places $= 4! = 24$.

Number of ways in which all the four objects can occupy their places $= 1$.

If three objects occupy their places, 4th will also do so. So this is contained in possibility discussed above.

If two objects occupy their places, remaining two can go wrong by occupying each other's position *i.e.*, in only one way.

Since out of 4, two objects can be chosen in $4c_2$ ways, number of ways in which only two objects can occupy their places.

$$= 4c_2 \times 1 = 6.$$

If one object occupies its position, any one of the remaining three can go wrong in 2 ways by occupying the positions of other two.

Since out of 4, one object can be chosen in $4c_1$ ways, number of ways in which only one object can occupy its place.

$$= 4c_1 \times 2 = 8.$$

$\therefore$ Total number of ways in which at least one object can occupy its place

$$= 1+6+8 = 15$$

$\therefore$ Total number of ways in which none of the objects occupies the place corresponding to its number

$$= 24 - 15 = 9$$

$$\therefore \text{ Reqd. prob. } = \frac{9}{24} = \frac{3}{8}.$$

Ex. 7-18. *A and B stand in a ring with 10 other persons. If the arrangement of 12 persons is at random, find the chance that there are exactly three persons between A and B.*

Sol. There are in all 12 persons who are to stand in a ring. Fixing the position of one person remaining 11 persons can stand in a ring $= 11!$

$\therefore$ Total number of ways in which 12 persons can stand in a ring $= 11!$

Out of 10 persons three are to stand between A and B. These three can be chosen in ${}^{10}c_3$ ways and can be arranged in 3! ways.

Also number of ways of arranging remaining 7 persons = 7!

Since A and B can interchange their position, number of ways of having exactly 3 persons between A and B

$$= 2.^{10}c_3\ 3!\ 7$$

$$= 2.\frac{10!}{3!7!}.3!.7$$

$$= 2.10!$$

$\therefore$ Reqd. prob.

$$= \frac{2.10!}{11!} = \frac{2}{11}.$$

Ex. 7-19. *The first 12 letters of the alphabet are written at random. Find the chance that there are exactly 4 letters between A and B.*

Sol. Total number of ways = 12!

Different possibilities are :

1	2	3	4	5	6	7	8	9	10	11	12
A	.	.	.	.	B	.	.	.	.	.	.
.	A	.	.	.	.	B	.	.	.	.	.
.	.	A	.	.	.	.	B	.	.	.	.
.	.	.	A	.	.	.	.	B	.	.	.
.	.	.	.	A	.	.	.	.	B	.	.
.	.	.	.	.	A	.	.	.	.	B	.
.	.	.	.	.	.	A	.	.	.	.	B

Out of remaining 10 letters, 4 letters are to lie between A and B. These four can be chosen in $10c_4$ ways and can be arranged in 4! ways.

Also number of ways of arranging remaining 6 letters = 6! and A and B can interchange their positions.

$\therefore$ Total number of ways of having exactly 4 letters between A and B.

$$= 7.2.^{10}c_4.4!.6$$

$$= 7.2\cdot\frac{10!}{6!\cdot 4!}\cdot 4!.6$$

$$= 7.2.10!$$

$\therefore$ Reqd. prob. $\dfrac{7.2.10!}{12!} = \dfrac{7}{66}$.

Ex. 7-20. *If the letters of the word 'REGULATIONS' be arranged at random, what is the chance that there will be exactly 4 letters between the 'R' and the 'E' ?*

Sol. There are in all 11 letters to be arranged.

Total number of ways of arranging 11 letters

$$= 11!$$

Different possibilities are :

1	2	3	4	5	6	7	8	9	10	11
R	.	.	.	.	E	.	.	.	.	.
.	R	.	.	.	.	E	.	.	.	.
.	.	R	.	.	.	.	E	.	.	.
.	.	.	R	.	.	.	.	E	.	.
.	.	.	.	R	.	.	.	.	E	.
.	.	.	.	.	R	.	.	.	.	E

Therefore, as in last examp
the *'R' and the 'E'*

Therefore, reqd. prob.

Ex. 7-21. *Show that the ch*

Sol. There are six faces of
$\therefore$ Total number of cases =
Out of six faces, three are
$\therefore$ Number of favourable c

$\therefore$ Reqd. prob.

Ex. 7-22. *In a single throw*
eleven.

Sol. Total number of cases
(*i*) The sum 8 can be obtain
Fi

$\therefore$ The number of favourab

$\therefore$ Reqd. prob.

(*ii*) The sum 11 can be obta
Fi

$\therefore$ Number of favourable c

$\therefore$ Reqd. probability

Ex. 7-23. *Find the chance*
Sol. Consider the expressio

Therefore, as in last example, total number of ways of having exactly 4 letters between the '*R*' *and the* '*E*'

$$= 6.2.^{9}c_4.4!.5$$
$$= 6.2.9!$$

Therefore, reqd. prob. $= \dfrac{6.2.9!}{11!}$

$$= \frac{6.2}{11.10} = \frac{6}{55}.$$

Ex. 7-21. *Show that the chance of throwing an odd number with a die is* $\frac{1}{2}$.

Sol. There are six faces of a die marked with numbers from 1 to 6.

$\therefore$ Total number of cases = 6.

Out of six faces, three are marked with odd numbers *viz*., 1, 3 and 5.

$\therefore$ Number of favourable cases = 3.

$\therefore$ Reqd. prob. $= \dfrac{3}{6}$

$$= \frac{1}{2}.$$

Ex. 7-22. *In a single throw with two dice, find the chances of throwing (i) eight, (ii) eleven.*

Sol. Total number of cases = 6 × 6 = 36.

(*i*) The sum 8 can be obtained in either of the following ways :

First die	Second die
6	2
5	3
4	4
3	5
2	6

$\therefore$ The number of favourable cases = 5.

$\therefore$ Reqd. prob. $= \dfrac{5}{36}.$

(*ii*) The sum 11 can be obtained in either of the following ways :

First die	Second die
6	5
5	6

$\therefore$ Number of favourable cases = 2.

$\therefore$ Reqd. probability $= \dfrac{2}{36} = \dfrac{1}{18}.$

Ex. 7-23. *Find the chance of throwing a total of 3 or 5 or 11 with two dice.*

Sol. Consider the expression

$$(x + x^2 + ... + x^6)^2$$

$$= \left\{\frac{x(1-x^6)}{1-x}\right\}^2$$

$$= x^2(1-x^6)^2(1-x)^{-2}$$

$$= x^2(1-2x^6+x^{12})(1+2x+3x^2+4x^3+....)$$

$\therefore$ Number of ways of getting a total of 3

$$= \text{co-efficient of } x^3 = 2$$

Number ot ways of getting a total of 5

$$= \text{co-efficient of } x^5 = 4$$

and Number of ways of getting a total of 11

$$= \text{co-efficient of } x^{11} = 10-8 = 2$$

$\therefore$ Number of ways of getting a total of 3 or 5 or 11

$$= 2+4+2 = 8$$

Total no. of cases $= 6^2 = 36$

$\therefore$ Reqd. prob. $= \dfrac{8}{36} = \dfrac{2}{9}.$

Ex. 7-24. *Show that the chance of throwing 15 with 3 dice is* $\dfrac{5}{108}$.

Sol. Consider the expression

$$(x+x^2+...+x^6)^3$$

$$= x^3\frac{(1-x^6)^3}{(1-x)^3}$$

$$= x^3(1-3x^6+3x^{12}-x^{18})(1-x)^{-3}$$

$$= x^3(1-3x^6+3x^{12}-x^{18})\left(1+3x+\frac{3.4}{2!}x^2+\frac{3.4.5}{3!}x^3+.........+\frac{3.4.5...(r+2)}{r!}x^r+......\right)$$

Therefore, number of ways of getting a total of 15

$$= \text{co-efficient of } x^{15}$$

$$= \frac{3.4.5.6.7.8.9.10.11.12.13.14}{12!} - 3.\frac{3.4.5.6.7.8}{6!} + 3$$

$$= 10.$$

Total number of cases $= 6^3 = 216$

$\therefore$ Reqd. prob. $= \dfrac{10}{216} = \dfrac{5}{108}.$

Ex. 7-25. *Show that, in a single throw with two dice, the chance of throwing more than 7 is equal to that of throwing less than 7 each being* $\dfrac{5}{12}$.

Sol. Consider the expression

$$(x+x^2+....+x^6)^2$$

$$= x^2(1-x^6)^2(1-x)^{-2}$$

$$= x^2(1-2x^6+x^{12})(1+2x+3x^2+4x^3+...+(r+1)x^r+...)$$

$= x^2(1+2x+3x^2+4x^3+5$

Therefore, number of ways

and number of ways of getti

Total number of cases

Therefore, probability of ge

Ex. 7-26. *Each co-efficient i an ordinary die. Find the prob. t*

Sol. The roots of the given e

Now various possible values

Since there are 3 co-efficient

Now various possibilities are

ac	*a*	*c*
1	1	1
2	2	1
	1	2
3	1	3
	3	1
4	1	4
	2	2
	4	1
5	1	5
	5	1
6	1	6
	2	3
	3	2
	6	1
7	1	7
	7	1
8	2	4
	4	2
9	3	3

(Values 10, 11 etc, of '*ac*' are not

$+3x^2+4x^3+\ldots)$

$-8=2$

is $\frac{5}{108}$.

$)(1-x)^{-3}$

$.+\frac{3.4.5\ldots(r+2)}{r!}x^r+\ldots\ldots\Big)$

$\frac{.14}{}-3.\frac{3.4.5.6.7.8}{6!}+3$

hance of throwing more than

$$=x^2(1+2x+3x^2+4x^3+5x^4+6x^5+5x^6+4x^7+3x^8+2x^9+x^{10}+\ldots)$$

Therefore, number of ways of getting a number less than 7

$$= \text{sum of co-efficients of } x^2, x^3, x^4, x^5 \text{ and } x^6$$
$$=1+2+3+4+5=15$$

and number of ways of getting a number greater than 7

$$=5+4+3+2+1=15$$

Total number of cases $= 6^2 = 36$

Therefore, probability of getting a number greater than 7

$$= \text{probability of getting a number less than 7}$$
$$=\frac{15}{36}=\frac{5}{12}.$$

Ex. 7-26. *Each co-efficient in the equation* $ax^2+bx+c=0$ *is determined by throwing an ordinary die. Find the prob. that the equation will have real roots.*

Sol. The roots of the given *equation will be real if* $b^2 \geq 4ac$.

Now various possible values of co-efficients *in the* equation are 1, 2, 3, 4, 5, 6.
Since there are 3 co-efficients, total number of cases

$$=6^3=216.$$

Now various possibilities are :

ac	*a*	*c*	4*ac*	*b*	*No. of cases*
1	1	1	4	2, 3, 4, 5, 6	$5\times1=5$
2	2 1	1 2	8	3, 4, 5, 6	$4\times2=8$
3	1 3	3 1	12	4, 5, 6	$3\times2=6$
4	1 2 4	4 2 1	16	4, 5, 6	$3\times3=9$
5	1 5	5 1	20	5, 6	$2\times2=4$
6	1 2 3 6	6 3 2 1	24	5, 6	$2\times4=8$
7	1 7	7 1	This is not possible as on a die number greater than '6' can't occur.		
8	2 4	4 2	32	6	$1\times2=2$
9	3	3	36	6	$1\times1=1$

(Values 10, 11 etc, of '*ac*' are not possible as $b^2 \not> 36$).

$\therefore$ Total number of favourable cases

$$= 43$$

$$\therefore \text{ Reqd. prob.} \qquad = \frac{43}{216}.$$

Ex. 7-27. *Four tickets marked 00, 01, 10, 11 respectively are placed in a bag. A ticket is drawn at random five times, being replaced each time. Find the prob that the sum of the numbers on tickets thus drawn is 23.*

Sol. Number of favourable cases = co-efficient of x^{23} *in*

$$(x^0 + x^1 + x^{10} + x^{11})^5 \text{ i.e., } (1 + x + x^{10} + x^{11})^5$$

i.e., $\quad (1+x)^5 (1+x^{10})^5$

i.e., $\quad (1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5)(1 + 5x^{10} + 10x^{20} + 10x^{30} + 5x^{40} + x^{50})$

$\therefore$ Number of favourable cases = 100

Total number of ways of drawing cards = 4^5

$$\therefore \text{ Reqd. prob.} \qquad = \frac{100}{4^5} = \frac{25}{256}.$$

Ex. 7-28. *An urn contains a white balls and b black balls. If $a + \beta$ balls are drawn from this urn, find the probability that among them there will be exactly α white and β black balls.*

Sol. Total number of cases = $^{a+b}c_{\alpha+\beta}$.

Number of ways of drawing α white balls

$$= {}^a c_\alpha$$

and number of ways of drawing β black balls

$$= {}^b c_\beta$$

$\therefore$ Number of ways of having α white and β black balls among $(\alpha + \beta)$ balls drawn

$$= {}^a c_\alpha . {}^b c_\beta$$

$$\therefore \text{ Reqd. prob.} \qquad = \frac{{}^a c_\alpha \cdot {}^b c_\beta}{{}^{a+b} c_{\alpha+\beta}}.$$

Ex. 7-29. *A die is cast until a 6 appears. Find the probability that it must be cast more than 5 times.*

$$\textbf{Sol. } \text{Reqd. prob.} \qquad = \left(\frac{5}{6}\right)^5 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^6 \cdot \frac{1}{6} + \ldots\ldots\ldots$$

$$= (5/6)^5 \cdot \frac{1}{6} \left\{1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \ldots\right\}$$

$$= \left(\frac{5}{6}\right)^5 \cdot \frac{1}{6} \cdot \frac{1}{1 - 5/6}$$

7.2. Notations

(*1*) *Capital letters A, B, C*

(*2*) (*A*), (*B*) *etc., denote t*

(*3*) $\overline{(A)}, \overline{(B)}$ *etc., denote*

(*4*) (*AB*) *denotes the simu*

(*5*) (*A+B*) *denotes the ha*

(*6*) *P*(*A*), *P*(*B*) *etc., deno*

(*7*) *P*(*A/B*) *denotes the co*

is known that the event B has

7.3. Theorem of Total Proba

It states that the probabili

events is the sum of the proba

$P(A_1 + A_2$

where $A_1, A_2 \ldots\ldots A_n$ *are M.E.*

Sol. Let $A_1, A_2, \ldots\ldots\ldots A_n$

$P(A_1 + A_2$

Let *N* be the number of cas

Out of these let

no. of cases favourable to

no. of cases favourable to

....

....

no. of cases favourable to

Since $A_1, A_2, \ldots\ldots\ldots A_n$ ar

distinct and no-overlapping.

$\therefore$ No. of cases which are

of the events $A_1, A_2, \ldots\ldots\ldots\ldots A$

$\therefore \qquad P(A_1 + A_2 + \ldots.$

Now by def.,

]

$\therefore \; P(A_1 + A_2 + \ldots + A_n) =$

$$= \left(\frac{5}{6}\right)^5$$

7.2. Notations

(1) Capital letters A, B, C etc., denote events.

(2) (A), (B) etc., denote the happening of events A, B etc.

(3) $\overline{(A)}, \overline{(B)}$ etc., denote the non-happenings of A, B etc.

(4) (AB) denotes the simultaneous happening of A and B.

(5) (A+B) denotes the happening of at least one of the events A and B.

(6) P(A), P(B) etc., denote the probabilities of the happening of the events A, B etc.

(7) P(A/B) denotes the conditional probability of the happening of the event A when it is known that the event B has already happened.

7.3. Theorem of Total Probability

It states that the probability of the happening of any one of the several mutually exclusive events is the sum of the probabilities of the happening of separate events i.e.,

$$P(A_1 + A_2 + \ldots\ldots + A_n) = P(A_1) + P(A_2) + \ldots\ldots + P(A_n)$$

where $A_1, A_2 \ldots\ldots A_n$ are M.E. events.

Sol. Let $A_1, A_2, \ldots\ldots\ldots A_n$ be n mutually exclusive events. Then we are to show that

$$P(A_1 + A_2 + \ldots\ldots + A_n) = P(A_1) + P(A_2) + \ldots\ldots + P(A_n)$$

Let N be the number of cases which are equally likely, mutually exclusive and exhaustive. Out of these let

no. of cases favourable to $A_1 = m_1$

no. of cases favourable to $A_2 = m_2$

....

....

no. of cases favourable to $A_n = m_n$

Since $A_1, A_2, \ldots\ldots\ldots A_n$ are mutually exclusive, the cases $m_1, m_2, \ldots\ldots\ldots m_n$ are quite distinct and no-overlapping.

$\therefore$ No. of cases which are favourable to $(A_1 + A_2 + \ldots\ldots\ldots + A_n)$ (*i.e.*, occurrence of any of the events $A_1, A_2, \ldots\ldots\ldots A_n$)

$$= m_1 + m_2 + \ldots + m_n$$

$$\therefore \quad P(A_1 + A_2 + \ldots\ldots + A_n) = \frac{m_1 + m_2 + \ldots\ldots + m_n}{N}$$

$$= \frac{m_1}{N} + \frac{m_2}{N} + \ldots\ldots + \frac{m_n}{N}$$

Now by def.,

$$P(A_1) = \frac{m_1}{N} \text{ etc.}$$

$$\therefore \quad P(A_1 + A_2 + \ldots\ldots + A_n) = P(A_1) + P(A_2) + \ldots\ldots + P(A_n)$$

7.3.1. Generalization of the theorem of total probability for non-mutually exclusive events.

Let $A_1, A_2 ....A_n$ be the events which are not mutually exclusive. Consider A_1 and A_2 two events. Two mutually exclusive and exhaustive forms in which A_1 can happen are :

(*i*) A_1 happens and A_2 does not happen *i.e.*, $(A_1\bar{A}_2)$

(*ii*) A_1 happens and A_2 also happens *i.e.*, (A_1A_2)

Let m_1 and m_2 be the number of cases favourable to $(A_1\bar{A}_2)$ and (A_1A_2) respectively.

Then number of cases favourable to $(A_1) = m_1 + m_2$

$$\therefore \qquad P(A_1) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n}$$

where n is the number of equally likely, mutually exclusive and exhaustive cases.

$$\therefore \qquad P(A_1) = P(A_1\bar{A}_2) + P(A_1A_2) \qquad ...(1)$$

Interchanging A_1 and A_2,

$$P(A_2) = P(\bar{A}_1A_2) + P(A_1A_2) \qquad ...(2)$$

Now the three mutually exclusive and exhaustive forms in which $(A_1 + A_2)$ can happen are

(*i*) A_1 happens and A_2 does not happen *i.e.*, $(A_1\bar{A}_2)$.

(*ii*) A_1 happens and A_2 also happens *i.e.*, (A_1A_2).

(*iii*) A_2 happens and A_1 does not happen *i.e.*, $(\bar{A}_1A_2)$.

$\therefore$ As before, $P(A_1 + A_2) = P(A_1\bar{A}_2) + P(A_1A_2) + P(\bar{A}_1A_2)$ ...(3)

Subtracting (1) and (2) from (3)

$$P(A_1 + A_2) - P(A_1) - P(A_2) = -P(A_1A_2)$$

$$\therefore \qquad P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1A_2) \qquad ...(4)$$

$$\begin{aligned}
\text{Now} \qquad P(A_1 + A_2 + A_3) &= P(A_1 + \overline{A_2 + A_3}) \\
&= P(A_1) + P(A_2 + A_3) - P(A_1\overline{A_2 + A_3}) \\
&= P(A_1) + P(A_2 + A_3) - P(A_1A_2 + A_1A_3) \\
&= P(A_1) + \{P(A_2) + P(A_3) - P(A_2A_3)\} \\
&\quad - \{P(A_1A_2) + P(A_1A_3) - P(A_1A_2A_3)\} \\
&\quad \{\because (A_1A_2A_1A_3) \equiv (A_1A_2A_3)\} \\
&= \sum_{i=1}^{3} P(A_i) - \sum_{\substack{i,j=1 \\ i<j}}^{3} P(A_iA_j) + (-1)^{3-1} P(A_1A_2A_3)
\end{aligned}$$

ty for non-mutually exclusive

y exclusive. Consider A_1 and

s in which A_1 can happen are

$_1\bar{A}_2)$ and (A_1A_2) respectively.

and exhaustive cases.

...(1)

...(2)

in which (A_1+A_2) can happen

$A_2)$ ...(3)

$_2)$

$_2)$ ...(4)

$-P(A_1\overline{A_2+A_3})$

$-P(A_1A_2+A_1A_3)$

$(A_3)-P(A_2A_3)\}$

$_3)-P(A_1A_2A_3)\}$

$)\equiv(A_1A_2A_3)\}$

$_iA_j)+(-1)^{3-1}P(A_1A_2A_3)$

∴ In general,

$$P(A_1+A_2+...+A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,j=1 \\ i<j}}^{n} P(A_iA_j)$$

$$\sum_{\substack{i<j<k=1 \\ i<j<k}}^{n} P(A_iA_jA_k)........+(-1)^{n-1}P(A_1A_2...A_n)$$

Remark 1 : If $A_1, A_2, ...A_n$ are mutually exclusive, $P(A_iA_j)=0$, $P(A_iA_jA_k)=0$ etc.

∴ $P(A_1+A_2+...+A_n) = P(A_1)+P(A_2)+...+P(A_n)$

Remark 2 : (4) is called additive law of probability for two events. It can also be proved without using theorem of total probability as seen below :

7.3.2. Additive Law of probability for the two events.

Let A_1 and A_2 be two events (not necessarily mutually exclusive). Let m_1, m_2 and m_3 be the number of cases favourable to $(A_1), (A_2)$ and (A_1A_2) respectively. Then

Number of cases favourable to $(A_1\bar{A}_2) = m_1 - m_3$

and number of cases favourable to $(\bar{A}_1A_2) = m_2 - m_3$

∴ Number of cases favourable to (A_1+A_2) are

$$(m_1-m_3)+(m_2-m_3)+m_3$$

i.e., $$m_1+m_2-m_3.$$

∴ If n be the total number of cases which are mutually exclusive and equally likely then

$$P(A_1+A_2) = \frac{m_1+m_2-m_3}{n}$$

$$= \frac{m_1}{n}+\frac{m_2}{n}-\frac{m_3}{n}$$

$$= P(A_1)+P(A_2)-P(A_1A_2)$$

Ex. 7-30. *Show that* $P(\varphi)=0$

Sol. In axiom (3) take

$$A_1 = A_2 = ... = \varphi$$

Then $$A_1+A_2+...A_n+... = \varphi$$

∴ Axiom (3) ⇒ $P(\varphi) = P(\varphi)+P(\varphi)+........$

⇒ $P(\varphi)=0$.

Remark : Taking $A_{n+1}=\varphi, A_{n+2}=\varphi, ...$

Axiom (3) ⇒

$P(A_1+A_2+....A_n) = P(A_1)+....+P(A_n)$ for any positive integer n.

7.3.3. *Using axiomatic approach show that*

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Sol. We have

$$A = AB \cup A\overline{B}$$

and $AB \cap A\overline{B} = \varphi$

$\therefore$ By axiom (3)

$$P(A) = P(AB) + P(A\overline{B})$$

$$\Rightarrow P(A\overline{B}) = P(A) - P(AB) \qquad \ldots(1)$$

Also $\quad A \cup B = B \cup A\overline{B}$

and $\quad B \cap A\overline{B} = \varphi$

$$\therefore \quad P(A \cup B) = P(B) + P(A\overline{B})$$

$$= P(B) + P(A) - P(AB).$$

Ex. 7-31. *Show that* $P(\overline{A}) = 1 - P(A)$

Sol. We have

$$A \cup \overline{A} = S \quad \text{and} \quad A \cap \overline{A} = \varphi$$

$$\therefore \quad P(S) = P(A) + P(\overline{A})$$

$$\Rightarrow \quad 1 = P(A) + P(\overline{A})$$

$$\Rightarrow \quad P(\overline{A}) = 1 - P(A)$$

Ex. 7-32. *If* $P(A) = \frac{1}{3}$, $P(\overline{B}) = \frac{1}{4}$, *can A and B be disjoint. Explain :*

Sol. Let if possible,

A and B be disjoint

$$\Rightarrow A \cap B = \varphi$$

$$\Rightarrow A \subseteq \overline{B}$$

$$\Rightarrow P(A) \le P(\overline{B})$$

$$\Rightarrow \frac{1}{3} \le \frac{1}{4} \text{ i.e., } 4 \le 3$$

which is not true.

$\therefore$ *A, B* can't be disjoint.

Ex. 7-33. *Prove or disprove :*

If $P(A) = P(\overline{B})$ *then* $\overline{A} = B$

Sol. Consider sample space

$$S = \{e_1, e_2, e_3, e_4, e_5\}$$

Where $P(e_i) = 0{\cdot}2 \quad (i = 1, 2 \ldots\ldots 5)$

Let $A = (e_2)$, $B = \{e_1, e_2, e_3, e_5\}$

Then $\overline{B} = \{e_4\}$

$\therefore$ $P(A) = P(\overline{B}) = 0{\cdot}2$

We have $\overline{A} = \{e_1, e_3, e_4,$

Ex. 7-34. *Prove or dispro*

If $P(A) = 0$, $P(AB) = 0$

Sol. $AB \subseteq A$

$\Rightarrow P(AB) \le P(A) = 0$

$\Rightarrow P(AB) = 0$

Ex. 7-35. *Prove or dispro*

If $P(A) = P(B) = p$, *then*

Sol. Consider a sample sp

where $P(e_1) = 0.15$, $P(e_2)$

$P(e_4) = 0.2 = P(e_5$

Consider $A = \{e_1, e_3, e_4\}$,

$P(A) = 0{\cdot}15 + 0{\cdot}1 + 0{\cdot}2$

$P(B) = 0{\cdot}15 + 0{\cdot}1 + 0{\cdot}2$

$\therefore$ $p = P(A) = P(B) = 0{\cdot}$

$AB = \{e_1, e_3\}$

$\therefore$ $P(AB) = 0{\cdot}15 + 0{\cdot}1 =$

Ex. 7-36. *Out of a group*

different birthdays (assume 36

Sol. Req.

where (3

Ex. 7-37. *If* $A \subset B$ *then*

Sol. Since $A \subset B$, $AB = A$

Also $B = BA \cup B\overline{A}$

$= A \cup B\overline{A}$

$\Rightarrow P(B) = P(A) + P(B\overline{A})$

$\Rightarrow P(B) \ge P(A)$

Ex. 7-38. *If* $P(\overline{A}) = \alpha$ *and*

$P(AB) \ge 1 - \alpha$

Sol. We have

$P(A$

)

...(1)

oint. Explain :

$\therefore\ P(A) = P(\overline{B}) = 0{\cdot}2$

We have $\overline{A} = \{e_1, e_3, e_4, e_5\} \neq B = \{e_1, e_2, e_3, e_5\}$.

Ex. 7-34. *Prove or disprove :*

If $P(A) = 0$, $P(AB) = 0$

Sol. $AB \subseteq A$

$\Rightarrow P(AB) \leq P(A) = 0$

$\Rightarrow P(AB) = 0$

Ex. 7-35. *Prove or disprove :*

If $P(A) = P(B) = p$, *then* $P(AB) \leq p^2$

Sol. Consider a sample space

$$S = \{e_1, e_2, e_3, e_4, e_5\}$$

where $P(e_1) = 0.15$, $P(e_2) = 0.35$, $P(e_3) = 0{\cdot}1$

$P(e_4) = 0.2 = P(e_5)$

Consider $A = \{e_1, e_3, e_4\}$, $B = \{e_1, e_3, e_5\}$

$P(A) = 0{\cdot}15 + 0{\cdot}1 + 0{\cdot}2 = 0{\cdot}45$

$P(B) = 0{\cdot}15 + 0{\cdot}1 + 0{\cdot}2 = 0{\cdot}45$

$\therefore\ p = P(A) = P(B) = 0{\cdot}45$

$AB = \{e_1, e_3\}$

$\therefore\ P(AB) = 0{\cdot}15 + 0{\cdot}1 = 0{\cdot}25 \nleq (0{\cdot}45)^2 = 0{\cdot}2025.$

Ex. 7-36. *Out of a group of 25 persons, what is the probability that all 25 will have different birthdays (assume 365 days in a year and all days equally likely).*

Sol. $$\text{Req. prob.} = \frac{(365)_{25}}{(365)^{25}}$$

where $(365)_{25} = (365).(365-1)\ldots\ldots(365-\overline{25-1})$.

Ex. 7-37. *If* $A \subset B$ *then* $P(A) \leq P(B)$

Sol. Since $A \subset B$, $AB = A$

Also $B = BA \cup B\overline{A}$

$= AUB\overline{A}$

$\Rightarrow P(B) = P(A) + P(B\overline{A}) \quad (\because A \cap B\overline{A} = \varphi)$

$\Rightarrow P(B) \geq P(A) \quad (\because P(B\overline{A}) \geq 0)$

Ex. 7-38. *If* $P(\overline{A}) = \alpha$ *and* $P(\overline{B}) = \beta$ *then*

$$P(AB) \geq 1 - \alpha - \beta.$$

Sol. We have

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(AB) \\ &= \{1 - P(\overline{A})\} + \{1 - P(\overline{B})\} - P(AB) \\ &= 2 - \alpha - \beta - P(AB) \end{aligned}$$

Now $\quad P(A \cup B) \le 1$

$\therefore \quad 2 - \alpha - \beta - P(AB) \le 1$

$\Rightarrow P(AB) \ge 1 - \alpha - \beta.$

Ex. 7-39. *If $AB = \varphi$, show that*

$$P(A) \le P(\overline{B}).$$

Sol. We have

$$P(A) = P(AB) + P(A\overline{B})$$

Since $\quad AB = \varphi,\ P(AB) = 0$

$$\therefore \quad P(A) = P(A\overline{B}) = P(\overline{B})P(A/\overline{B})$$
$$\le P(\overline{B}) \qquad \{\because P(A/\overline{B}) \le 1\}.$$

7.4. Independent Events Def.

Two events are said to be independent (in probability sense) if the probability of happening of one does not depend on the happening or non-happening of the other.

7.4.1. Theorem of Compound Probability.

If states that the probability of the simultaneous occurrence of two non-mutually exclusive events is equal to the prbability of hapening of one multiplied by the conditional probability of the other when it is known that first has already happened

i.e.,
$$P(AB) = \begin{cases} P(A)\,P(B/A) \\ \quad \text{or} \\ P(B)\,P(A/B) \end{cases}$$

Proof. Let A and B be two non-mutually exclusive events.

Two mutually exclusive and exhaustive forms in which A can happen are:

(1) A happens and B does not happen *i.e.,* $(A\overline{B})$.

(2) A happens and B also happens *i.e.,* (AB).

Let m_1 and m_2 be the number of cases favourable to $(A\overline{B})$ and (AB) respectively and n the number of cases which are equally likely, mutually exclusive and exhaustive.

Then number of cases favourable to $(A) = m_1 + m_2$

$$\therefore \quad P(A) = \frac{m_1 + m_2}{n}$$

Now by def.,
$$P(AB) = \frac{m_2}{n}$$
$$= \frac{m_1 + m_2}{n} \cdot \frac{m_2}{m_1 + m_2} = P(A) \cdot \frac{m_2}{m_1 + m_2}$$

Assuming the occurrence of A, out of n only $(m_1 + m_2)$ cases are left, out of which m_2 are also favourable to B.

$\therefore \quad \dfrac{m_2}{m_1 + m_2}$ gives the conditional probability of B when it is given that A has occurred

i.e.,

$\therefore$

Interchanging A and B

Remark 1. If A and B ar
the happening or non-happer

$\therefore \quad P(B/A) =$

$\therefore \quad P(AB) = P($

Converse of this also
independent.

(2) Some authors use "
instead of independence.

Ex. 7-40. *If A and B are*

(*i*) *A and* $\overline{B}$

Sol. Since A and B are i

$P(AB) = P(A$

(*i*) Now

$\Rightarrow A, \overline{B}$ are independen

Similarly (*ii*), (*iii*), can b

Ex. 7-41. *If* $P(B) > 0$, *s*

$P(\varphi / B) = 0$

Proof. $P(\varphi / B) = \dfrac{P(\varphi B}{P(B)}$

$= \dfrac{P(\varphi)}{P(B)}$

Ex. 7-42. *If $P(B) > 0$, sh*

$P(\overline{A} /$

Sol. $\quad P(\overline{A} /$

Also $\quad P($

$\Rightarrow P(\overline{A}B)$

$\therefore \quad (1) \quad \Rightarrow \quad P($

y sense) if the probability of
-happening of the other.

e of two non-mutually exclusive
l by the conditional probability
d

its.
A can happen are:

$4\bar{B}$) and (AB) respectively and
:lusive and exhaustive.

$= P(A) \cdot \dfrac{m_2}{m_1 + m_2}$

cases are left, out of which m_2

n it is given that A has occurred

$$\frac{m_2}{m_1 + m_2} = P(B / A)$$

$$\therefore \qquad P(AB) = P(A) P(B / A)$$

Interchanging A and B

$$P(AB) = P(B)\, P(A / B)$$

Remark 1. If A and B are independent, prob of happening of B (or A) is not affected by the happening or non-happening of A (or B)

$$\therefore \qquad P(B / A) = P(B) \text{ and } P(A / B) = P(A)$$

$$\therefore \qquad P(AB) = P(A) P(B)$$

Converse of this also holds *i.e.*, if this condition is satisfied events A and B are independent.

(2) Some authors use "statistically independence" or "stochastically independence" instead of independence.

Ex. 7-40. *If A and B are independent then so are*

(*i*) *A and $\bar{B}$* (*ii*) *$\bar{A}$ and B* (*iii*) *$\bar{A}$ and $\bar{B}$.*

Sol. Since A and B are independent, we have

$$P(AB) = P(A)\, P(B)$$

(*i*) Now

$$\begin{aligned} P(A\bar{B}) &= P(A) - P(AB) \\ &= P(A) - P(A)P(B) \\ &= P(A)\{1 - P(B)\} \\ &= P(A)P(\bar{B}) \end{aligned}$$

$\Rightarrow A, \bar{B}$ are independent

Similarly (*ii*), (*iii*), can be proved.

Ex. 7-41. *If $P(B) > 0$, show that*

$$P(\varphi / B) = 0$$

Proof. $P(\varphi / B) = \dfrac{P(\varphi B)}{P(B)}$

$$= \frac{P(\varphi)}{P(B)} = 0 \qquad (\because P(\varphi) = 0)$$

Ex. 7-42. *If $P(B) > 0$, show that*

$$P(\bar{A} / B) = 1 - P(A / B)$$

Sol.

$$P(\bar{A} / B) = \frac{P(\bar{A}B)}{P(B)} \qquad \ldots(1)$$

Also

$$P(B) = P(AB) + P(\bar{A}B)$$

$$\Rightarrow P(\bar{A}B) = P(B) - P(AB)$$

$$\therefore \quad (1) \quad \Rightarrow \quad P(\bar{A} / B) = 1 - \frac{P(AB)}{P(B)}$$

$$= 1 - P(A / B)$$

Ex. 7-43. *Prove or disprove: pairwise independence of events does not imply independence.*

Sol. Consider an experiment of tossing of two cubical dice.

Let A_1: odd face on first die

A_2: odd face on second die

A_3: sum of numbers on faces of two dice is odd.

$$P(A_1) = \frac{3}{6} = \frac{1}{2}, P(A_2) = \frac{1}{2}$$

For A_3, possibilities are :

(1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5)
(3, 2), (3, 4), (3, 6), (4, 1), (4, 3), (4, 5)
(5, 2), (5, 4), (5, 6), (6, 1), (6, 3), (6, 5)

$$\therefore \ P(A_3) = \frac{18}{36} = \frac{1}{2}$$

For $A_1 \cap A_2$, possibilities are :

(1, 1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5)
(5, 1), (5, 3), (5, 5)

$$\therefore \ P(A_1 \cap A_2) = \frac{9}{36} = \frac{1}{4} = P(A_1)\,P(A_2)$$

For $A_1 \cap A_3$, possibilities are :

(1, 2), (1, 4), (1, 6), (3, 2), (3, 4), (3, 6), (5, 2), (5, 4), (5, 6)

$$P(A_1 \cap A_3) = \frac{9}{36} = \frac{1}{4} = P(A_1)\,P(A_3)$$

Similarly,

$$P(A_2 \cap A_3) = \frac{9}{36} = \frac{1}{4} = P(A_2)P(A_3)$$

$\therefore (A_1; A_2); (A_1; A_3); (A_2; A_3)$ are independent.

For $A_1 \cap A_2 \cap A_3$ there is no possibility.

$$\therefore \ P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1)P(A_2)P(A_3)$$

7.4.2. If $P(B) > 0$ and $A_1, A_2 \ldots\ldots\ldots A_n$ are M.E. *events, show that*

$$P(A_1 \cup A_2 \cup \ldots \cup A_n / B) = P(A_1 / B) + \ldots + P(A_n / B)$$

Sol.
$$P(A_1 \cup \ldots \cup A_n / B) = \frac{p\{(A_1 \cup \ldots \cup A_n)B\}}{P(B)}$$
$$= \frac{P(A_1 B \cup \ldots\ldots \cup A_n B)}{P(B)}$$
$$= \sum_{i=1}^{n} \frac{P(A_i B)}{P(B)} \qquad \{\because A_i B\text{'s are also M.E.}\}$$

7.4.3. *If P (B) > 0 and*

$P(A_1 \cup A_2 / B) = P(A_1 /$

Sol. $P(A_1 \cup$

Ex. 7.44. *Given* $P(A) >$

show that

Sol. We have

and

$\therefore$

$\Rightarrow$

7.4.4. *If* $B_1, B_2, \ldots\ldots\ldots$

and $P(B_j) > 0$

then for any event A,

Proof. We have

$\therefore$

of events does not imply

e.

(5, 4), (5, 6)

, show that

$P(A_n / B)$

$\{\because A_i B$'s are also M.E.$\}$

$$= \sum_i P(A_i / B).$$

7.4.3. *If $P(B) > 0$ and A_1, A_2 any two events show that*

$P(A_1 \cup A_2 / B) = P(A_1 / B) + P(A_2 / B) - P(A_1 A_2 / B)$

Sol.
$$P(A_1 \cup A_2 / B) = \frac{P\{(A_1 \cup A_2)B\}}{P(B)}$$

$$\frac{P(A_1 B \cup A_2 B)}{P(B)}$$

$$= \frac{P(A_1 B) + P(A_2 B) - P(A_1 A_2 B)}{P(B)}$$

$$= \frac{P(A_1 B)}{P(B)} + \frac{P(A_2 B)}{P(B)} - \frac{P(A_1 A_2 B)}{P(B)}$$

$$= P(A_1 / B) + P(A_2 / B) - P(A_1 A_2 / B)$$

Ex. 7.44. *Given* $P(A) > 0, P(B) > 0$ *and*

$$P(A / B) = P(B / A)$$

show that
$$P(A) = P(B).$$

Sol. We have

$$P(A / B) = \frac{P(AB)}{P(B)}$$

and
$$P(B / A) = \frac{P(AB)}{P(A)}$$

$\therefore$
$$\frac{P(AB)}{P(A)} = \frac{P(AB)}{P(B)}$$

$\Rightarrow$
$$P(A) = P(B)$$

7.4.4. *If* $B_1, B_2, \ldots\ldots\ldots B_n$ are *M.E. events s.t.*

$$S = \bigcup_j B_j$$

and $\quad P(B_j) > 0 \qquad (j = 1, 2 \ldots\ldots n)$

then for any event A,

$$P(A) = \sum_j P(A / B_j) P(B_j)$$

Proof. We have

$$S = \bigcup_j B_j$$

$\therefore$
$$A = AS$$

$$= A \bigcup_j B_j$$

$$= \bigcup_j AB_j$$

Also AB_j's are M.E.

$$\therefore \quad P(A) = \sum_j P(AB_j)$$

$$= \sum_j P(A/B_j)P(B_j).$$

Remark : This theorem remains true if n is infinite.

Cor : Since

$$B \cup \overline{B} = S \quad \text{and} \quad P(B) > 0, P(\overline{B}) > 0$$

we have

$$P(A) = P(A/B)P(B) + P(A/\overline{B})P(\overline{B}).$$

Ex. 7-45. *Under what conditions does the following equality hold :*

$$P(A) = P(A/B) + P(A/\overline{B})$$

Sol. We have

$$P(A) = P(A/B)P(B) + P(A/\overline{B})P(\overline{B})$$

and by given

$$P(A) = P(A/B) + P(A/\overline{B})$$

$\therefore$ Subtracting

$$\therefore \quad 0 = P(A/B)\{1 - P(B)\} + P(A/\overline{B})\{1 - P(\overline{B})\} \qquad \ldots(1)$$

since $1 > P(B) > 0$ and $1 > P(\overline{B}) > 0$

we have

$1 - P(B) > 0$ and $1 - P(\overline{B}) > 0$

$\therefore$ (1) $\Rightarrow$

$$P(A/B) = 0 = P(A/\overline{B})$$

$\Rightarrow P(AB) = 0$ and $P(A\overline{B}) = 0$

$\Rightarrow P(A) = P(AB) + P(A\overline{B}) = 0$

$$\Rightarrow A = \varphi.$$

Ex. 7-46. *If A and B are two events and the probability* $P(B) \neq 1$, *prove that*

$$P(A/\overline{B}) = \{P(A) - P(AB)\}/\{1 - P(B)\}$$

Hence show that

$$P(AB) \geq P(A) + P(B) - 1.$$

Sol. By compound prob. theorem

$$P(A\overline{B}) = P(\overline{B})P(A/\overline{B})$$

$$\Rightarrow \quad P(A/\overline{B}) = \frac{P(A\overline{B})}{P(\overline{B})} \qquad \ldots(1)$$

Now $\quad P(A) = P(AB) + P(A\overline{B})$

$\Rightarrow \quad P(A\overline{B}) = P(A) - (AB)$

Also $\quad P(\overline{B}) = 1 - P(B)$

$$\therefore \qquad (1) \Rightarrow \therefore P(A/\bar{B}) = \frac{P(A) - P(AB)}{1 - P(B)} \qquad \ldots(2)$$

$$\text{Now } P(A/\bar{B}) \le 1$$

$$\therefore \qquad (2) \Rightarrow P(A) - P(AB) \le 1 - P(B)$$

$$\Rightarrow P(AB) \ge P(A) + P(B) - 1.$$

Ex. 7-47. *Three fair dice are thrown once. Given that no two show the same face. Find the probability that*

(*i*) *the sum of the faces is 7*

(*ii*) *one face is an ace.*

Sol. Let E : No two die show the same face

S : Sum of the faces.

$$P(E) = \frac{6.54}{6.6.6} = \frac{5}{9}$$

(*i*) To find $P(S = 7 \cap E)$:

Different possibilities are :

(1, 2, 4); (1, 4, 2); (2, 1, 4); (2, 4, 1); (4, 1, 2); (4, 2, 1)

$$\therefore \qquad P(S = 7 \cap E) = \frac{6}{216} = \frac{1}{36}$$

$$S(S = 7/E) = \frac{P(S = 7 \cap E)}{P(E)} = \frac{1/36}{5/9} = \frac{1}{20}.$$

(*ii*) Let A : one face is an ace.

To find $P(A \cap E)$:

If ace comes on first die, different possibilities are :

(1, 2, 3); (1, 3, 4); (1, 4, 5); (1, 5, 6)

(1, 3, 2); (1, 4, 3); (1, 5, 4); (1, 6, 5)

But there are three dice and ace can occur on any dice.

$\therefore$ No. of possibilities $= 3 \times 8 = 24$.

$$\therefore \; P(A \cap E) = \frac{24}{216} = \frac{1}{9}$$

$$\therefore \; P(A/E) = \frac{P(A \cap E)}{P(E)} = \frac{\frac{1}{9}}{5/9} = \frac{1}{5}.$$

Ex. 7-48. *A die is thrown as long as necessary for a 6 to turn up. Given that 6 does not turn up at the first throw, what is the probability that more than four throws will be necessary ?*

Sol. Let E_1 : Event that 6 does not turn at the first throw.

E_2 : more than four throws are necessary (*i.e.*, 6 does not turn up in first four throws).

$$P(E_1) = {}^5\!/_6.$$

$P(E_1 E_2)$ = Prob. that 6 does not turn up in first four throws.

$$= \left(\tfrac{5}{6}\right)^4$$

$$P(E_2 / E_1) = \frac{P(E_1 E_2)}{P(E_1)} = \left(\tfrac{5}{6}\right)^3 = \frac{125}{216}.$$

Ex. 7-49. *Given* $P(A) = 0 \cdot 5$ *and* $P(A \cup B) = 0 \cdot 6$, *find* $P(B)$ *if :*

(*i*) $A; B$ are mutually exclusive.
(*ii*) $A; B$ are independent.
(*iii*) $P(A / B) = 0 \cdot 4$.

Sol. (*i*) $P(A \cap B) = 0$ as A, B are M.E.

$$\therefore \quad P(A \cup B) = P(A) + P(B)$$

$$\Rightarrow \quad 0{\cdot}6 = 0{\cdot}5 + P(B)$$

$$P(B) = 0 \cdot 1$$

(*ii*)
$$P(A \cap B) = P(A)P(B), \text{ as } A; B \text{ are independent}$$
$$= (0{\cdot}5)\, P(B)$$

Also
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$0{\cdot}6 = 0 \cdot 5 + P(B)\{1 - 0 \cdot 5\}$$
$$0{\cdot}1 = P(B)(0 \cdot 5)$$
$$P(B) = \frac{0 \cdot 1}{0 \cdot 5} = \frac{1}{5} = 0 \cdot 2$$

(*iii*)
$$P(A \cap B) = P(B)P(A / B) = 0 \cdot 4P(B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$0{\cdot}6 = 0 \cdot 5 + P(B)\{1 - 0 \cdot 4\}$$
$$\therefore \quad P(B) = \frac{0 \cdot 1}{0 \cdot 6} = 1/6.$$

Ex. 7-50. *If A and B are independent and* $P(A) = P(B / A) = \frac{1}{2}$, *find* $P(A \cup B)$.

Sol. Since A and B are independent,

$$P(B) = P(B / A) = \frac{1}{2}$$

$$P(AB) = P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$\therefore \quad P(A \cup B) = P(A) + P(B) - P(AB)$$
$$= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \tfrac{3}{4}.$$

Ex. 7-51. *If* $P(A) = P(B) = P(B / A) = \frac{1}{2}$, *are A, B independent ?*

Sol. $P($

$\therefore$ A, B are independent.

Ex. 7-52. *If A and B are in*

Sol. $P(A\overline{B} \cup$

Ex. 7-53. *If* $P(B) = P(A / B$

Sol. $P($

$P(A$

Ex. 7-54. *Suppose* B_1, B_2, B

$(j = 1, 2, 3)$. *Find* $P(A)$.

Sol. $P(AB$

As B's are *M.E.*,

$P($

Ex. 7-55. *Prove the followin*

(*i*) *if* $P(A / B) \geq P(A)$ *then*

(*ii*) *if* $P(B / \overline{A}) = P(B / A)$,

(*iii*) *if* $P(A) = a$, $P(B) = b$,

Sol.

$$P(AB) = P(A)P(B/A)$$

$$= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} = P(A)P(B)$$

$\therefore$ A, B are independent.

Ex. 7-52. *If A and B are independent and* $P(A) = P(B) = \frac{1}{2}$, *find* $P(A\overline{B} \cup \overline{A}B)$.

Sol.

$$P(A\overline{B} \cup \overline{A}B) = P(A\overline{B}) + P(\overline{A}B)$$

$(\because A\overline{B}, \overline{A}B$ are $M.E.)$

$$= P(A)P(\overline{B}) + P(\overline{A})P(B)$$

$$= \frac{1}{2}\times\left(1-\frac{1}{2}\right)+\left(1-\frac{1}{2}\right)\times\frac{1}{2}$$

$$= \frac{1}{2}.$$

Ex. 7-53. *If* $P(B) = P(A/B) = P(C/AB) = \frac{1}{2}$, *find* $P(ABC)$.

Sol.

$$P(AB) = P(B)P(A/B) = \frac{1}{4}$$

$$P(ABC) = P(AB)P(C/AB)$$

$$= \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8}.$$

Ex. 7-54. *Suppose* B_1, B_2, B_3 *are mutually exclusive. If* $P(B_j) = \frac{1}{3}$ *and* $P(A/B_j) = \frac{j}{6}$, $(j = 1, 2, 3)$. *Find* $P(A)$.

Sol.

$$P(AB_j) = P(B_j)P(A/B_j)$$

$$= \frac{1}{3}\cdot\frac{j}{6} = \frac{j}{18}.$$

As B's are $M.E.$,

$$P(A) = P(AB_1) + P(AB_2) + P(AB_3)$$

$$= \frac{1}{18}+\frac{2}{18}+\frac{3}{18} = \frac{1}{3}.$$

Ex. 7-55. *Prove the following :*

(*i*) *if* $P(A/B) \ge P(A)$ *then* $P(B/A) \ge P(B)$

(*ii*) *if* $P(B/\overline{A}) = P(B/A)$, *then A and B are independent*

(*iii*) *if* $P(A) = a$, $P(B) = b$, *then*

$$P(A/B) \ge \frac{a+b-1}{b}.$$

Sol. (*i*) $$P(A/B)P(B) = P(AB) = P(A)P(B/A)$$

$\therefore$ $$\frac{P(A/B)}{P(A)} = \frac{P(B/A)}{P(B)} \geq 1$$

$\Rightarrow$ $$P(B/A) \geq P(B)$$

(*ii*) **By given**

$$P(B/\bar{A}) = P(B/A)$$

$\Rightarrow$ $$\frac{P(B\bar{A})}{P(\bar{A})} = \frac{P(AB)}{P(A)}$$

$\Rightarrow$ $$P(A)P(B\bar{A}) = P(AB)\{1-P(A)\}$$

$\Rightarrow$ $$P(A)\{P(B\bar{A})+P(AB)\} = P(AB)$$

$\Rightarrow$ $$P(A)\,P(B) = P(AB)$$

$$\{\because P(B\bar{A})+P(AB) = P(B)\}$$

$\therefore$ *A*, *B* are independent.

(*iii*) By additive law,

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

also $$P(A \cup B) \leq 1$$

$\therefore$ $$P(A) + P(B) - P(AB) \leq 1.$$

$$P(AB) \geq a+b-1$$

$\Rightarrow$ $$P(B)\,P(A/B) \geq a+b-1$$

$\Rightarrow$ $$P(A/B) \geq \frac{a+b-1}{b}.$$

Ex. 7-56. *A single die is tossed; then n coins are tossed, where n is the number on the die. Find the probability of exactly two heads.*

Sol. Possible values of *n* are :

$$2, 3, 4, 5, 6$$

and probability of each is $\frac{1}{6}$.

$$\therefore \text{ Reqd. prob.} = \frac{1}{6}\left\{{}^2c_2\left(\frac{1}{2}\right)^2 + {}^3c_2\left(\frac{1}{2}\right)^3 + {}^4c_2\left(\frac{1}{2}\right)^4 + {}^5c_2\left(\frac{1}{2}\right)^5 + {}^6c_2\left(\frac{1}{2}\right)^6\right\}$$

$$= \frac{1}{6}\left\{\frac{1}{4} + \frac{3}{8} + \frac{3}{8} + \frac{5}{16} + \frac{15}{64}\right\} = \frac{33}{128}.$$

Ex. 7-57. *Prove or disprove the following :*

if $P(A) > P(B)$ *then* $P(A/C) > P(B/C)$.

Sol. Consider biased tetrahedron with faces *a, b, c* and *d* with respective probabilities

$$0{\cdot}1,\ 0{\cdot}2,\ 0{\cdot}2 \text{ and } 0{\cdot}5$$

Define $$A = \{a, b, d\}$$

$$B = \{b, c\}, \quad C = \{a, b, c\}$$

l)

$\{\because P(B\overline{A})+P(AB)=P(B)\}$

?)

', *where n is the number on the*

$\left.{}^5c_2\left(\frac{1}{2}\right)^5+{}^6c_2\left(\frac{1}{2}\right)^6\right\}$

$\left.+\frac{15}{64}\right\}=\frac{33}{128}.$

d with respective **probabilities**

then
$$P(A) = 0\cdot1+0\cdot2+0\cdot5=0\cdot8$$
$$P(B) = 0\cdot2+0\cdot2=0\cdot4$$
$$\therefore \quad P(A)>P(B)$$
and
$$P(C) = 0\cdot1+0\cdot2+0\cdot2=0\cdot5$$
Now
$$A\cap C = \{a,b\}\Rightarrow P(A\cap C)=0\cdot1+0\cdot2=0\cdot3$$
$$B\cap C = \{b,c\}\Rightarrow P(B\cap C)=0\cdot2+0\cdot2=0\cdot4$$
$$P(A/C) = \frac{P(A\cap C)}{P(C)}=\frac{0\cdot3}{0\cdot5}=0\cdot6$$
$$P(B/C) = \frac{P(B\cap C)}{P(C)}=\frac{0\cdot4}{0\cdot5}=0\cdot8$$
$$\therefore \quad P(A/C)\not> P(B/C)$$

Ex. 7-58. *Let* $B_1, B_2,\ldots\ldots B_n$ *be mutually disjoint and let* $B=\bigcup_{j=1}^{n}B_j$. *Suppose* $P(B_j)>0$ *and* $P(A/B_j)=p,\ j=1,2\ldots\ldots n$. *Show that* $P(A/B)=p$.

Sol. Let
$$P(B_j) = p_j>0$$
Then
$$P(B) = \sum_{j=1}^{n}P(B_j)=\sum_{j=1}^{n}p_j$$
Also
$$AB = \bigcup_{f=1}^{n}AB_j \text{ and } AB_j \text{ are mutually disjoint.}$$
$$\therefore \quad P(AB) = \sum_{j=1}^{n}P(AB_j).$$
$$= \sum_j P(B_j)P(A/B_j)$$
$$= \sum_j p_jp.=p\left(\sum_j p_j\right)$$
$$\therefore \quad P(A/B) = \frac{P(AB)}{P(B)}=\frac{p\left(\sum_j p_j\right)}{\left(\sum_j p_j\right)}=p.$$

Ex. 7-59. *It is given that*

$$P(A_1 + A_2) = \frac{5}{6},\ P(A_1A_2) = \frac{1}{3},\ P(\overline{A}_2) = \frac{1}{2}$$

where $P(\overline{A}_2)$ *stand for the probability that* A_2 *does not happen. Determine* $P(A_1)$ *and* $P(A_2)$. *Hence show that the events* A_1 *and* A_2 *are independent.*

Sol. Since total probability is always unity,

$$P(A_2) + P(\overline{A}_2) = 1$$

$$\therefore \quad P(A_2) = 1 - P(\overline{A}_2)$$

$$= 1 - \frac{1}{2} = \frac{1}{2}.$$

By additive law for non-mutually exclusive events,

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1A_2)$$

$$\therefore \quad \frac{5}{6} = P(A_1) + \frac{1}{2} - \frac{1}{3}$$

$$\therefore \quad P(A_1) = \frac{2}{3}$$

$$\therefore \quad P(A_1)P(A_2) = \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3} = P(A_1A_2)$$

$\therefore$ Events A_1 and A_2 are independent.

Ex. 7-60. *Discuss and criticise the following :*

$$P(A) = \frac{2}{3},\ P(B) = \frac{1}{4},\ P(C) = \frac{1}{6}$$

where A, B and C are mutually exclusive events.

Sol. Since three events *A, B* and *C* are mutually exclusive, by total probability theorem

$$P(A+B+C) = P(A) + P(B) + P(C)$$

$$= \frac{2}{3} + \frac{1}{4} + \frac{1}{6}$$

$$= \frac{13}{12} > 1$$

Since the probability is always less than unity, the statement is wrong.

Ex. 7-61. *Two packs of cards are made up in such a way that the first pack consists of 39 red cards and 13 black cards; second pack consists of 39 black cards and 13 red cards. A sampling experiment is carried out in the following way: A card is drawn from the first pack, if it is red, a second card is drawn from the same pack after replacing the first red card. The colour of the second card drawn from the first pack is noted. If the first card drawn from the first pack is black, then the second card is drawn from the second pack and the colour of the second card is noted. Both the cards are then replaced in their respective packs. What is the probability that the second card is red ?*

Sol. Two different possibilities are :

(1) First card drawn is red.

(2) First card drawn is black.

(1) Now the probability

Since the first card drawı
but after replacing the first re

∴ Probability of drawin

This is the conditional p
known that card drawn in firs
∴ By the theorem of com
and second draws.

(2) The probability of dr

Since the first card drawn
∴ Conditional probabilit
drawn in first draw.

∴ Probability of drawing

∴ Since two possibilities

Req. prob.

Ex. 7-62. *From each of th*
What is the probability of thei
Sol. Probability of selecti

There are only two mutua
(1) All the partners are of
(2) All the partners are of
By compound probability

∴ By theorem of total pr

(1) Now the probability of drawing a red card from the first pack.

$$= \frac{39}{39+13}$$

$$= \frac{39}{52} = \frac{3}{4}.$$

Since the first card drawn is red, the second card is also to be drawn from the first pack but after replacing the first red card.

$\therefore$ Probability of drawing a red card in second draw $= \frac{3}{4}$.

This is the conditional probability of drawing a red card in second draw when it is known that card drawn in first draw was red.

$\therefore$ By the theorem of compound probability, the probability of drawing red cards in first and second draws.

$$= \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}.$$

(2) The probability of drawing a black card from the first pack

$$= \frac{13}{39+13} = \frac{13}{52} = \frac{1}{4}$$

Since the first card drawn is black, the second card is to be drawn from the second pack.

$\therefore$ Conditional probability of drawing a red card when it is known that a black card was drawn in first draw.

$$= \frac{13}{39+13} = \frac{1}{4}$$

$\therefore$ Probability of drawing a black card in first draw and a red card in second draw

$$= \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}$$

$\therefore$ Since two possibilities are mutually exclusive, by the theorem of total probability.

Req. prob. $$= \frac{9}{16} + \frac{1}{16} = \frac{10}{16} = \frac{5}{8}.$$

Ex. 7-62. *From each of three married couples one of the partners is selected at random. What is the probability of their being all of one sex ?*

Sol. Probability of selecting a partner (male or female) from either couple

$$= \frac{1}{2}$$

There are only two mutually exclusive possibilities :

(1) All the partners are of male sex,

(2) All the partners are of female sex.

By compound probability theorem, probability of either possibility

$$= \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$\therefore$ By theorem of total probability,

Reqd. prob. $= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

Ex. 7-63. *In above question, show that the probability of choosing two men and one woman is* $\frac{3}{8}$.

Sol. Three possibilities are :

1st couple	*2nd couple*	*3rd couple*
M	*M*	*W*
M	*W*	*M*
W	*M*	*M*

Probability of choosing a partner (male or female) from either couple

$$= \frac{1}{2}$$

By compound probability theorem, probability of either possibility

$$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

$$= \frac{1}{8}.$$

By theorem of total probability, probability of choosing two men and one woman

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

$$= \frac{3}{8}.$$

Ex. 7-64. *A number x is chosen at random from the integers 1, 2, 3, ...n, and A and B denote the event that x is a multiple of 2 and 3 respectively. Show that A and B are independent events when n = 96 but not when n = 100.*

Sol. The no. of integers in 1, 2,....n, which are divisible by the integer m is the greatest integer less than $\frac{n}{m}$

(*i*) When $n = 96$

No. of integers which are divisible by 2

$$= \frac{96}{2} = 48$$

No. of integers which are divisible by 3

$$= \frac{96}{3} = 32$$

No. of integers which are divisible by 3 and 2 both *i.e.*, by 6

$$= \frac{96}{6} = 16$$

$\therefore$ $P(A) =$ Probability that x is a multiple of 2

$P(B) =$ Probability that x

and P(AB) = Probability tl

$\therefore$ $P(AB) = P(A)P(B)$

$\therefore$ Events A and B are inde

(*ii*) When

No. of integers which are c

No. of integers which are c

No. of integers which are d

$\therefore$

and P(

$\therefore$ P(

$\therefore$ Events A and B are not i

Ex. 7-65. *A bag contains 5 ... replaced and then a second dra... drawn were of different colours*

Sol. The two different poss

(1) The first draw gives wh

(2) The first draw gives bla

Since the ball drawn in the 1 (white or black) in two draws a

Now the probability of dra

and the probability of drawing a

$$= \frac{48}{96} = \frac{1}{2}$$

$P(B) =$ Probability that x is a multiple of 3

$$= \frac{32}{96} = \frac{1}{3}$$

and P(AB) = Probability that x is a multiple of 2 and 3 both *i.e.*, 6

$$= \frac{16}{96} = \frac{1}{6}$$

$\therefore\ P(AB) = P(A)P(B)$

$\therefore$ Events A and B are independent.

(*ii*) When $n = 100$

No. of integers which are divisible by 2

$$= \frac{100}{2} = 50$$

No. of integers which are divisible by 3

$$= \text{greatest integer less than } \frac{100}{3} = 33$$

No. of integers which are divisible by 3 and 2 both *i.e.*, 6

$$= \text{greatest integer less than } \frac{100}{6} = 16$$

$$\therefore \qquad P(A) = \frac{50}{100} = \frac{1}{2}$$

$$P(B) = \frac{33}{100}$$

and
$$P(AB) = \frac{16}{100}$$

$$\therefore \qquad P(AB) \neq P(A)P(B)$$

$\therefore$ Events A and B are not independent.

Ex. 7-65. *A bag contains 5 white and 4 black balls. A ball is drawn from this bag and replaced and then a second draw of a ball is made. What is the probability that the two balls drawn were of different colours ?*

Sol. The two different possibilities are :

(1) The first draw gives white ball and the second draw gives black ball.

(2) The first draw gives black ball and the second draw gives white ball.

Since the ball drawn in the first draw is replaced, the probabilities of drawing either ball (white or black) in two draws are same.

Now the probability of drawing a white ball

$$= \frac{5}{9}$$

and the probability of drawing a black ball

$$= \frac{4}{9}$$

∴ By the theorem of compound probability, probability of possibility (1)

$$= \frac{5}{9} \times \frac{4}{9} = \frac{20}{81}$$

and probability of possibility (2)

$$= \frac{4}{9} \times \frac{5}{9} = \frac{20}{81}$$

∴ By the theorem of total probability, the probability of getting two balls of different colours in two draws

$$= \frac{20}{81} + \frac{20}{81} = \frac{40}{81}.$$

Ex. 7-66. *An urn contains 4 white and 5 black balls, a second urn contains 5 white and 4 black balls. One ball is transferred from the first to second urn, then a ball is drawn from the second urn. What is the probability that it is white ?*

Sol. There are two different possibilities.

(1) The ball transferred from the first to second urn is white.

(2) The ball transferred from the first to second urn is black.

The probability of drawing a white ball from second urn in these two possibilities will be different. So we consider these two possibilities separately.

(1) Now probability of drawing a white ball from the first urn

$$= \frac{4_{c_1}}{9_{c_1}} = \frac{4}{9}$$

Since the ball transferred from first to second urn is white, total number of white balls in second run

$$= 5 + 1 = 6.$$

The number of balls in second urn

$$= 6 + 4 = 10$$

∴ Probability of drawing a white ball from second urn

$$= \frac{6}{10} = \frac{3}{5}$$

∴ By the theorem of compound probability, the probability of transferring a white ball and then drawing a white ball from the second urn

$$= \frac{4}{9} \cdot \frac{3}{5} = \frac{4}{15}$$

(2) **Probability** of drawing a black ball from the first urn

$$= \frac{5_{c_1}}{9_{c_1}} = \frac{5}{9}$$

Total number of white balls in second urn = 5 and number of balls in second urn = 10.

∴ Probability of drawing a white ball from the second urn

$$= \frac{5}{10} = \frac{1}{2}$$

∴ The probability of transferring a black ball and then drawing a white ball from the second urn

of possibility (1)

getting two balls of different

ond urn contains 5 white and
urn, then a ball is drawn from

ite.
ack.
in these two possibilities will
r.
st urn

, total number of white balls in

ity of transferring a white ball

er of balls in second urn = 10.
urn

drawing a white ball from the

$$= \frac{5}{9} \cdot \frac{1}{2} = \frac{5}{18}$$

The two possibilities (1) and (2) are mutually exclusive. Hence by the theorem of total probability, the probability of transferring a ball from first urn to second and then drawing a white ball from the second.

$$= \frac{4}{15} + \frac{5}{18} = \frac{49}{90}$$

Ex. 7-67. *Three urns contain respectively 1 white, 2 black balls, 2 white and 1 black balls, 2 white and 2 black balls. One ball is transferred from the first urn into the second; then one from the latter is transferred into the third. Finally one ball is drawn from the third urn. What is the probability of its being white ?*

Sol. There are in all four different possibilities :

(1) The white ball is transferred from the first to the second urn and then a white ball is transferred from the second to the third urn.

(2) The white ball is transferred from the first to the second urn and then the black ball is transferred from the second to the third urn.

(3) A black ball is transferred from the first to the second urn and then a white ball is transferred from the second to the third urn.

(4) A black ball is transferred from the first to the second urn and then a black ball is transferred from the second to the third urn.

(1) Probability of drawing the white ball from the first urn

$$= \frac{1}{3}$$

After transferring the white ball from the first to the second urn,

Number of white balls in the second urn = 3

and total number of balls in the second urn = 4

∴ Probability of drawing a white ball from the second urn after transferring the white ball from the first to the second urn

$$= \frac{3}{4}$$

After transferring a white ball from the second to the third urn,

Number of white ball in the third urn = 3

and total number of balls in the third urn = 5

∴ The probability of drawing a white ball from the third urn after transferring a white ball from the second to the third urn

$$= \frac{3}{5}$$

∴ By the theorem of compound probability, the probability of transferring the white ball from the first to the second urn; then a white ball from the second to the third and then drawing a white ball from the third urn

$$= \frac{1}{3} \cdot \frac{3}{4} \cdot \frac{3}{5} = \frac{9}{60}$$

(2) After transferring white ball from the first to the second urn, the number of black balls in the second urn = 1.

∴ The probability of drawing the black ball from the second urn after transferring the white ball from the first to the second urn.

$$= \frac{1}{4}$$

After transferring the black ball from the second to the third urn, the number of white balls in the third urn = 2

$\therefore$ Probability of drawing a white ball from the third urn after transferring the black ball from the second to the third urn

$$= \frac{2}{5}.$$

$\therefore$ The probability of transferring the white ball from the first to the second; then the black ball from the second to the third and then drawing a white ball from the third urn

$$= \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{2}{60}.$$

(3) Proceeding as in above two cases, the probability of transferring a black ball from the first to the second urn; then a white ball from the second to the third urn and then drawing a white ball from the third urn

$$= \frac{2}{3} \cdot \frac{2}{4} \cdot \frac{3}{5} = \frac{12}{60}.$$

(4) The probability of transferring a black ball from the first to the second; a black ball from the second to the third and then drawing a white ball from the third urn

$$= \frac{2}{3} \cdot \frac{2}{4} \cdot \frac{2}{5} = \frac{8}{60}.$$

Since the four possibilities are mutually exclusive, by the theorem of total probability, the probability of transferring a ball from the first to the second; then a ball from the second to the third and then drawing a white ball from the third urn

$$= \frac{9}{60} + \frac{2}{60} + \frac{12}{60} + \frac{8}{60} = \frac{31}{60}.$$

Ex. 7-68. *In a bag there are six balls of which 3 are white and 3 are black. They are drawn successively without replacement. What is the chance that the colours are alternate ?*

Sol. Let (W) be the event that in a draw white ball appears and (B) be the event that black ball appears. The possible sequences are

$$(W)\,(B)\,(W)\,(B)\,(W)\,(B)$$

and

$$(B)\,(W)\,(B)\,(W)\,(B)\,(W)$$

Probability of drawing a white (or black) ball in first draw.

$$= \frac{3}{6} = \frac{1}{2}$$

Probability of drawing a black (or white) ball in second draw when the ball drawn in first draw is white (or black)

$$= \frac{3}{5}$$

Probability of drawing a white (or black) ball in third draw when in first two draws white (or black) and black (or white) balls have been drawn

$$= \frac{2}{4} = \frac{1}{2}$$

Probability of drawing a black (or white) ball in fourth draw when in third draw the ball drawn is white (or black)

$$= \frac{2}{3}$$

Probability of drawing a white (or black) ball in fifth draw

$$= \frac{1}{2}$$

and probability of drawing a black (or white) ball in sixth draw = 1.

$\therefore$ By compound probability theorem, probability of either sequence

$$= \left(\frac{1}{2}\right)\left(\frac{3}{5}\right)\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)(1)$$

$$= \frac{1}{20}$$

$\therefore$ By total probability theorem, probability of getting balls of alternate colours

$$= 2 \times \frac{1}{20} = \frac{1}{10}.$$

Ex. 7-69. *An urn contains 3 white and 5 black balls. One ball is drawn and its colour unnoted laid aside. Then another ball is drawn. Find the probability that it is white.*

Sol. There are two mutually exclusive possibilities :

(1) In the first draw a white ball appears.

Prob. of drawing a white ball in first draw

$$= \frac{3}{8}$$

Number of white balls in the urn before second draw

$$= 2.$$

Therefore, conditional probability of drawing a white ball in second draw, when in first draw a white ball has been drawn

$$= \frac{2}{7}$$

Therefore, by compound probability theorem, prob. of drawing a white ball in second draw

$$= \frac{3}{8} \cdot \frac{2}{7} = \frac{6}{56}$$

(2) In the first draw a black ball appears.

Prob. of drawing a black ball in first draw

$$= \frac{5}{8}$$

Number of white balls in the urn before second draw

$$= 3$$

Therefore, conditional prob. of drawing a white ball in second draw, when in first draw a black ball has been drawn

$$= \frac{3}{}$$

Therefore, prob. of drawing a white ball in second draw

$$= \frac{5}{8} \cdot \frac{3}{7} = \frac{15}{56}$$

$$\therefore \qquad \text{Reqd. prob.} = \frac{6}{56} + \frac{15}{56} = \frac{21}{56}.$$

Ex. 7-70. *A lady declares that by taking a cup of tea with milk she can discriminate whether the milk or tea-infusion was first added to the cup. It is proposed to test this assertion by means of an experiment with 10 cups of tea, five made in one way and five in the other and presenting them to lady for judgement in random order.*

Calculate the probability on the null hypothesis (i.e., the lady has no discrimination power) that the lady would judge correctly all the ten cups, being known to her 5 are of each kind.

Suppose that the tea cups were presented to the lady in five pairs, each pair to consist of cups of each kind in a random order. How would the probability of correctly judging with every cup on the null hypothesis be altered in this case ?

Sol. (*a*) When tea cups are presented in random order :

No. of ways of presenting 10 cups, 5 of each kind

$$= \frac{10!}{5!5!}$$

$$= 252.$$

Out of these 252 ways, the cups are presented in any one manner. The lady has to find which one is that method.

$\therefore$ No. of favourable cases = 1.

$$\therefore \qquad \text{Reqd. prob.} = \frac{1}{252}$$

(*b*) When cups are presented in 5 pairs :

Two ways of presenting a pair are :

$$M\,I; \quad I\,M$$

where 'M' stands for the cup prepared by taking milk first and 'I' for the cup prepared by taking infusion first.

When a pair is presented to the lady, she has to find, out of these two which one is the method used.

$\therefore$ Probability of correct judging for each pair

$$= \frac{1}{2}.$$

As the presentation of various pairs is independent of each other, by the theorem of compound probability, the joint probability of correctly judging the 5 pairs

$$= \left(\frac{1}{2}\right)^5 = \frac{1}{32}.$$

Ex. 7-71. *In a group of equal number of men and women 10% men and 45% women are unemployed. What is the probability that a person selected at random is employed ?*

Sol. Probability for a man to be unemployed

$$= \frac{10}{100} = \frac{1}{10}$$

and probability for a woman to

$\therefore$ Probability for a man to

and probability for a woman to

Since the group contains e

man $= \frac{1}{2}$ and same is probabili

$\therefore$ Probability of selecting

and probability of selecting an

The selected person may probability, probability of selec

Ex. 7-72. *In a random sam reading newspaper A and 400 that the habits of reading newsp chance that a person selected a many persons out of 1000 shou*

Sol. Probability for a perso

and probability for a person to b

Since the habits of reading theorem of compound probabilit reading both the newspapers

'th milk she can discriminate
proposed to test this assertion
one way and five in the other

e lady has no discrimination
'ng known to her 5 are of each

ive pairs, each pair to consist
ility of correctly judging with

: manner. The lady has to find

nd '*I*' for the cup prepared by

: of these two which one is the

each other, by the theorem of
;ing the 5 pairs

'en 10% men and 45% women
ted at random is employed ?

and probability for a woman to be unemployed

$$= \frac{45}{100} = \frac{9}{20}$$

$\therefore$ Probability for a man to be employed

$$= 1 - \frac{1}{10} = \frac{9}{10}$$

and probability for a woman to be employed

$$= 1 - \frac{9}{20} = \frac{11}{20}.$$

Since the group contains equal number of men and women, probability of selecting a man $= \frac{1}{2}$ and same is probability of selected a woman.

$\therefore$ Probability of selecting an employed man

$$= \left(\frac{1}{2}\right)\left(\frac{9}{10}\right)$$

and probability of selecting an employed woman

$$= \left(\frac{1}{2}\right)\left(\frac{11}{20}\right)$$

The selected person may be either man or woman. Hence by the theorem of total probability, probability of selecting an employed person.

$$= \left(\frac{1}{2}\right)\left(\frac{9}{10}\right) + \left(\frac{1}{2}\right)\left(\frac{11}{20}\right)$$

$$= \frac{1}{2}\left\{\frac{9}{10} + \frac{11}{20}\right\} = \frac{29}{40}.$$

Ex. 7-72. *In a random sample of 1000 residents of a large city, 700 were found to be reading newspaper A and 400 were found to be reading newspaper B. On the hypothesis that the habits of reading newspapers A and B are independent of each other,* (*i*) *what is the chance that a person selected at random would be reading both the newspapers,* (*ii*) *How many persons out of 1000 should be expected to be reading both newspapers ?*

Sol. Probability for a person to be reading newspaper A

$$= \frac{700}{1000} = \frac{7}{10}$$

and probability for a person to be reading newspaper B

$$= \frac{400}{1000} = \frac{2}{5}.$$

Since the habits of reading newspapers A and B are independent of each other, by the theorem of compound probability, the probability that a person selected at random would be reading both the newspapers

$$= \left(\frac{7}{10}\right)\left(\frac{2}{5}\right) = \frac{7}{25}$$

$\therefore$ No. of persons out of 1000 to be reading both the newspapers

$$= \frac{7}{25} \times 1000 = 280.$$

Ex. 7-73. *The probability of n independent events are* $p_1, p_2, p_3, \ldots\ldots. p_n$. *Find an expression for probability that at least one of the events will happen.*

Sol. Since total prob. is unity,

Prob. of non-happening of 1st event $= 1 - p_1$

Prob. of non-happening of 2nd event $= 1 - p_2$

....

Prob. of non-happening of nth event $= 1 - p_n$

Since the events are independent, by compound prob. theorem, Prob. of non-happening of all the events

$$= (1-p_1)(1-p_2)\ldots(1-p_n)$$

$\therefore$ Prob. of happening of at least one of the events

$$= 1-(1-p_1)(1-p_2)\ldots(1-p_n)$$

Ex. 7-74. *A problem in statistics is given to students whose chances of solving it are* $\frac{1}{2}, \frac{1}{3}$ *and* $\frac{1}{4}$ *respectively. What is the probability that the problem will be solved ?*

Sol. The problem will be solved if at least one student solves it.

Here $\quad p_1 = \frac{1}{2}, p_2 = \frac{1}{3}, p_3 = \frac{1}{4}.$

$\therefore$ Reqd. prob. $\quad 1-\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{4}\right)$

$$= \frac{3}{4}.$$

Ex. 7-75. *Three groups of children contain 3 girls and 1 boy, 2 girls and 2 boys and 3 boys and 1 girl. One child is selected at random from each group. Show that the chance that the three selected consist of 1 girl and 2 boys is* $\frac{13}{32}$.

Sol. Probabilities of selecting a boy from three groups are $\frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$ respectively and probabilities of selecting a girl from three groups are $\frac{3}{4}, \frac{1}{2}$ and $\frac{1}{4}$ respectively.

Different mutually exclusive possibilities of required selection are :

1st group	*2nd group*	*3rd group*
1 girl	1 boy	1 boy
1 boy	1 girl	1 boy
1 boy	1 boy	1 girl

By theorem of compound probability, probabilities of these possibilities are

$$\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{3}{4} = \frac{9}{32}, \frac{1}{4}\cdot\frac{1}{2}\cdot\frac{3}{4} = \frac{3}{32} \text{ and } \frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{4} = \frac{1}{32}$$

respectively.

Therefore, by the theorem

Ex. 7-76. *Four persons a*

and 4 children. Show that the

Sol. Total number of pers

Total number of ways of

Since out of 4 persons ch

chosen out of 3 men and 2 wo

Number of ways of select

and number of ways of selecti

$\therefore$ Total number of ways

$\therefore$ Required Prob

Ex. 7-77. *A and B are two*

a problem correctly are $\frac{1}{8}$ *an*

same answer. If the probabilit

answer was correct.

Sol. Let (A_1) and (A_2) be

answer was correct.

Two students can get the s

(1) Both of them get the c

(2) Both of them get the w

Therefore, by total probab

$P(A_1) = P$ (both students

$+ P$ (both students

respectively.

Therefore, by the theorem of total probability, probability of selecting 1 girl and 2 boys

$$= \frac{9}{32} + \frac{3}{32} + \frac{1}{32} = \frac{13}{32}.$$

Ex. 7-76. *Four persons are chosen at random from a group containing 3 men, 2 women and 4 children. Show that the chance that exactly two of them will be children is* $\frac{10}{21}$.

Sol. Total number of persons in the group.

$$= 3 + 2 + 4 = 9$$

Total number of ways of selecting 4 persons out of 9

$$= {}^9c_4$$

Since out of 4 persons chosen exactly two are to be children, remaining two are to be chosen out of 3 men and 2 women *i.e.*, 5 persons.

Number of ways of selecting two children out of 4

$$= {}^4c_2$$

and number of ways of selecting two persons out of 5

$$= {}^5c_2$$

$\therefore$ Total number of ways of having exactly 2 children in the selection of 4

$$= {}^4c_2 \cdot {}^5c_2$$

$$\therefore \qquad \text{Required Probability} = \frac{{}^4c_2\ {}^5c_2}{{}^9c_4} = \frac{10}{21}$$

Ex. 7-77. *A and B are two very weak students of statistics and their chances of solving a problem correctly are* $\frac{1}{8}$ *and* $\frac{1}{12}$ *respectively. They are given a question and obtain the same answer. If the probability of a common mistake is* $\frac{1}{1001}$, *find the chance that their answer was correct.*

Sol. Let (A_1) and (A_2) be the events that two students get the same answer and their answer was correct.

Two students can get the same answer in following two mutually exclusive ways :

(1) Both of them get the correct answer.

(2) Both of them get the wrong answer committing a common mistake.

Therefore, by total probability theorem,

$P(A_1) = P$ (both students get the correct answer)

$\qquad + P$ (both students get the wrong answer committing a common mistake)

$$= \frac{1}{8} \cdot \frac{1}{12} + \left(1 - \frac{1}{8}\right)\left(1 - \frac{1}{12}\right) \cdot \frac{1}{1001}$$

$$= \frac{1}{96} + \frac{77}{96.1001}$$

$$= \frac{1078}{96.1001}$$

Also $$P(A_1A_2) = \frac{1}{8}\cdot\frac{1}{12}$$

$$= \frac{1}{96}$$

By compound probability theorem,

$$P(A_1A_2) = P(A_1)P(A_2/A_1)$$

Therefore, $$P(A_2/A_1) = \frac{P(A_1A_2)}{P(A_1)}$$

$$= \frac{1001}{1078} = \frac{13}{14}.$$

Ex. 7-78. *A speaks truth in 75% and B in 80% of the cases. In what percentages of cases are they likely to contradict each other in stating the same fact ?*

Sol. A and B can contradict each other in following mutually exclusive ways :

(*i*) A speaks truth and B does not,

(*ii*) B speaks truth and A does not.

Therefore, by theorem of total probability, probability that A and B contradict each other

$$= \frac{3}{4}\cdot\frac{1}{5}+\frac{1}{4}\cdot\frac{4}{5}$$

$$= \frac{7}{20}$$

$$\left\{\begin{array}{ll} \because \text{ Prob. of } A \text{ speaking truth} = \dfrac{75}{100} = \dfrac{3}{4} \\ \text{Prob. of } B \text{ speaking truth} = \dfrac{80}{100} = \dfrac{4}{5} \end{array}\right\}$$

Therefore, A and B are likely to contradict each other in $\frac{7}{20}\times 100 = 35\%$ of the cases.

Ex. 7-79. *The probability that a teacher will give an unannounced test during any class meeting is $\frac{1}{5}$. If a student is absent twice, what is the probability that he will miss at least one test ?*

Sol. The student will not miss any test, if on the two days he is absent, the teacher does not give any test.

Probability for a teacher not giving any test on the two days (when the student is absent)

$$= \left(1-\frac{1}{5}\right)\left(1-\frac{1}{5}\right) = \frac{16}{25}$$

$\therefore$ Probability that the student will not miss any test

Since total probability is unity, p

Ex. 7-80. *A can solve 75% of the that either A or B can solve a proble*

Sol. Let (A) and (B) be the even

Then $P(A)$

and $P(B)$

Now $P(A+B)$

Therefore, probability that eithe

Ex. 7-81. *A husband and wife a post. The probability of the husband' What is the probability that only one*

Sol. There are only two M.E. po

(1) Husband is selected and wife

(2) Wife is selected and husband

Since there are two vacancies, hu

$\therefore$ Reqd. prob. =

=

=

Ex. 7-82. *A and B throw altern before B throws 6 and B wins if he thr the events that A wins and B wins the A's and B's turn to throw the dice, sho*

$$= \frac{16}{25}$$

Since total probability is unity, probability that the student will miss at least one test

$$= 1 - \frac{16}{25} = \frac{9}{25}$$

Ex. 7-80. *A can solve 75% of the problems and B can solve 70%. What is the probability that either A or B can solve a problem chosen at random ?*

Sol. Let (A) and (B) be the events that A and B solve the problem respectively.

Then $$P(A) = \frac{3}{4}$$

and $$P(B) = \frac{7}{10}$$

Now
$$\begin{aligned} P(A+B) &= P(A) + P(B) - P(AB) \\ &= P(A) + P(B) - P(A)P(B) \\ &= \frac{3}{4} + \frac{7}{10} - \frac{21}{40} \\ &= \frac{37}{40} \end{aligned}$$

Therefore, probability that either A or B can solve the problem

$$= \frac{37}{40}.$$

Ex. 7-81. *A husband and wife appear in an interview for two vacancies in the same post. The probability of the husband's selection is $\frac{1}{7}$ and that of the wife's selection is $\frac{1}{5}$. What is the probability that only one of them will be selected ?*

Sol. There are only two M.E. possibilities :

(1) Husband is selected and wife is not selected,

(2) Wife is selected and husband is not selected.

Since there are two vacancies, husband's selection and wife's selection are independent.

$$\therefore \quad \text{Reqd. prob.} = \frac{1}{7}\left(1 - \frac{1}{5}\right) + \left(1 - \frac{1}{7}\right) \cdot \frac{1}{5}$$

$$= \frac{4}{35} + \frac{6}{35} = \frac{10}{35}$$

$$= \frac{2}{7}.$$

Ex. 7-82. *A and B throw alternately a pair of unbiased dice. A wins if he throws 7 before B throws 6 and B wins if he throws 6 before A throws 7. If A and B respectively denote the events that A wins and B wins the series, a and b respectively denote the events that it is A's and B's turn to throw the dice, show that*

(*i*) $$P(A/a) = \frac{1}{6} + \frac{5}{6}P(A/b)$$

(*ii*) $$P(A/b) = \frac{31}{36}P(A/a)$$

(*iii*) $$P(B/a) = \frac{5}{6}P(B/b)$$

(*iv*) $$P(B/b) = \frac{5}{36} + \frac{31}{36}P(B/a)$$

Hence or otherwise find P(A/a) and P(B/a). Also comment on the result that

$$P(A/a) + P(B/a) = 1.$$

Sol. Now with a pair of dice,

prob. or throwing $6 = \frac{5}{36}$

and prob. of throwing $7 = \frac{6}{36} = \frac{1}{6}$

(*i*) $P(A/a) =$ Prob. that A wins if he has to throw first.

This is possible in following two mutually exclusive ways :

(*a*) A throws 7 in its first trial.

Its prob. is $\frac{1}{6}$.

(*b*) A does not throw 7 in its first throw. Then B will have the turn of throw and hence prob. of A's winning is $P(A/b)$.

$\therefore$ Prob. for this possibility $= \left(1 - \frac{1}{6}\right)P(A/b)$

$$= \frac{5}{6}P(A/b)$$

$\therefore$ By total prob. theorem

$$P(A/a) = \frac{1}{6} + \frac{5}{6}P(A/b)$$

(*ii*) $P(A/b) =$ Prob. that A wins if B is to throw first.

This is possible only when B does not throw 6 in its first throw. Then A has the turn to throw. So prob of A's winning is $P(A/a)$.

$\therefore$ By compound prob. theorem

$$P(A/b) = \left(1 - \frac{5}{36}\right)P(A/a)$$

$$= \frac{31}{36}P(A/a)$$

Similarly others can be proved.

ient on the result that

ays :

ave the turn of throw and hence

irst throw. Then A has the turn to

Ex. 7-83. *If 4 whole numbers taken at random are multiplied together, show that the chance that the last digit in the product is 1, 3, 7 or 9 is* $\frac{16}{625}$.

Sol. In all there are ten digits *viz.*, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Digits 1, 3, 7 and 9 have the property that when any two of them are multiplied, the last digit in the product is one of these four digits.

Therefore, if the last digit in the product of four whole numbers is to be 1, 3, 7 or 9 each whole number must have its last digit 1, 3, 7 or 9.

Probability for whole number to have its last digit 1, 3, 7 or 9.

$$= \frac{4}{10} = \frac{2}{5}$$

Therefore, by compound probability theorem probability that all the four whole numbers have their last digit 1, 3, 7 or 9

$$= \left(\frac{2}{5}\right)^4$$

$$= \frac{16}{625}$$

Therefore, reqd. prob. $= \frac{16}{625}$.

Ex. 7-84. *A six faced die is so biased that it is twice as likely to show an even number as an odd number when thrown. It is thrown twice. What is the probability that the sum of the two numbers thrown is even ?*

Sol. Let p be the probability for an odd number.

Then $2p$ is the probability for an even number.

When the die is thrown, either even number turns up or odd number turns up.

$\therefore \quad 2p + p = \text{Total prob.} = 1.$

$\therefore \quad p = \frac{1}{3}$

$\therefore \quad$ Prob. for an odd number $= \frac{1}{3}$

Prob. for an even number $= \frac{2}{3}$.

The sum of the two numbers thrown will be even if in both throws either we get even numbers or odd numbers. Now probability that in two throws even number turns up

$$= \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{4}{9}$$

and probability that in two throws odd numbers turn up

$$= \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{9}.$$

$\therefore$ By the theorem of total probability, probability that the sum of the two numbers thrown is even

$$= \frac{4}{9} + \frac{1}{9} = \frac{5}{9}.$$

Ex. 7-85. *A and B throw with a pair of dice. A wins if he throws 6 before B throws 7 and B wins if he throws 7 before A throws 6. If A begins, show that his chance of winning is* $\frac{30}{61}$.

Sol. Let (A) and (B) be the events that A gets 6 and B gets 7 with a pair of dice respectively.

Then $$P(A) = \frac{5}{36} \text{ and } P(B) = \frac{6}{36} = \frac{1}{6}.$$

Therefore $$P(\bar{A}) = 1 - \frac{5}{36} = \frac{31}{36} \text{ and } P(\bar{B}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Since A begins, he can win in following mutually exclusive ways :

$$(A), (\bar{A}\bar{B}A)\,(\bar{A}\bar{B}\bar{A}\bar{B}A) + \ldots\ldots\ldots\ldots\ldots\ldots$$

Therefore, by theorem of total probability, probability that A wins

$$= P(A) + P(\bar{A}\bar{B}A) + P(\bar{A}\bar{B}\bar{A}\bar{B}A) + \ldots$$

Since throws are independent, by compound probability theorem, probability that A wins

$$= P(A) + P(\bar{A})P(\bar{B})P(A) + P(\bar{A})P(\bar{B})P(\bar{A})P(\bar{B})P(A) + \ldots$$

$$= \frac{5}{36}\left\{1 + \left(\frac{31}{36}\cdot\frac{5}{6}\right) + \left(\frac{31}{36}\cdot\frac{5}{6}\right)^2 + \ldots\ldots\right\}$$

$$= \frac{5}{36}\cdot\frac{1}{1 - \frac{31}{36}\cdot\frac{5}{6}}$$

$$= \frac{30}{61}.$$

Ex. 7-86. *A and B take turns in throwing two dice, the first to throw 9 being awarded the prize. Show that their chances of winning are in the ratio 9 : 8.*

Sol. Let (A) and (B) be two events that A and B get 9 in a throw respectively.

Then $$P(A) = P(B) = \frac{4}{36} = \frac{1}{9}$$

Therefore $$P(\bar{A}) = P(\bar{B}) = 1 - \frac{1}{9} = \frac{8}{9}.$$

Since A begins, he can win in following mutually exclusive ways :

$$(A), (\bar{A}\bar{B}A), (\bar{A}\bar{B}\bar{A}\bar{B}A).$$

Therefore, by theorem of total probability, probability that A wins

$$= P(A) + P(\bar{A}\bar{B}A) + P(\bar{A}\bar{B}\bar{A}\bar{B}A) + \ldots\ldots\ldots\ldots\ldots\ldots$$

$$= P(A) + P(\bar{A})P(\bar{B})P(A) + P(\bar{A})P(\bar{B})P(\bar{A})P(\bar{B})P(A) + \ldots\ldots.$$

$$= \frac{1}{9}\left\{1 + \left(\frac{8}{9}\right)^2 + \left(\frac{8}{9}\right)^4 + \ldots.\right\}$$

Since total probability is unit

Therefore, chances of winnin

Ex. 7-87. *A, B and C in orde*

their respective chances of winni

Sol. Let (A), (B) and (C) be t

Then $P($

$\therefore$ $P($

Since A begins, he can win in

$(A), (\bar{A}$

Therefore, by theorem of tota

$= P(A) + P(\bar{A}\bar{B}$

$= P(A) + P(\bar{A})P$

B can win in following mutua

$(\bar{A}B), (\bar{A}\bar{B}\bar{C}\bar{A}B), (\bar{A}\bar{B}\bar{C}\bar{A}\bar{B}\bar{C}$

$\therefore$ Probability that B wins

$= P(\bar{A}B) + P(\bar{A}\bar{B}\bar{C}\bar{A}B) + P($

$= P(\bar{A})P(B) + P(\bar{A})P(\bar{B})P(\bar{C}$

$\therefore$ Probability that C wins

$$= \frac{1}{9} \cdot \frac{1}{1 - \frac{64}{81}} = \frac{9}{17}.$$

Since total probability is unity and one of two players is to win, probability that B wins

$$= 1 - \frac{9}{17} = \frac{8}{17}.$$

Therefore, chances of winning are in the ratio 9 : 8.

Ex. 7-87. *A, B and C in order toss a coin. The first one to throw a head wins. What are their respective chances of winning assuming that the game may continue indefinitely ?*

Sol. Let (A), (B) and (C) be the events that A, B and C get head in a toss respectively.

Then $$P(A) = P(B) = P(C) = \frac{1}{2}$$

$\therefore$ $$P(\bar{A}) = P(\bar{B}) = P(\bar{C}) = 1 - \frac{1}{2} = \frac{1}{2}$$

Since A begins, he can win in following mutually exclusive ways :

$$(A), (\bar{A}\bar{B}\bar{C}A), (\bar{A}\bar{B}\bar{C}\bar{A}\bar{B}\bar{C}A), \ldots\ldots\ldots\ldots$$

Therefore, by theorem of total probability, probability that A wins

$$= P(A) + P(\bar{A}\bar{B}\bar{C}A) + P(\bar{A}\bar{B}\bar{C}\ \bar{A}\bar{B}\bar{C}A) + \ldots\ldots\ldots$$

$$= P(A) + P(\bar{A})P(\bar{B})P(\bar{C})P(A) + \{P(\bar{A})P(\bar{B})P(\bar{C})\}^2 P(A) + \ldots\ldots$$

$$= \frac{1}{2} + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^7 + \ldots\ldots$$

$$= \frac{1}{2}\left\{1 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^6 + \ldots\ldots\right\} = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{8}} = \frac{4}{7}$$

B can win in following mutually exclusive ways :

$$(\bar{A}B), (\bar{A}\bar{B}\bar{C}\bar{A}B), (\bar{A}\bar{B}\bar{C}\bar{A}\bar{B}\bar{C}\bar{A}B), \ldots\ldots\ldots$$

$\therefore$ Probability that B wins

$$= P(\bar{A}B) + P(\bar{A}\bar{B}\bar{C}\bar{A}B) + P(\bar{A}\bar{B}\bar{C}\bar{A}\bar{B}\bar{C}\bar{A}B) + \ldots\ldots$$

$$= P(\bar{A})P(B) + P(\bar{A})P(\bar{B})P(\bar{C})P(\bar{A})P(B) + \{P(\bar{A})P(\bar{B})P(\bar{C})\}^2 P(\bar{A})P(B) + \ldots$$

$$= \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^8 + \ldots\ldots\ldots$$

$$= \frac{\frac{1}{4}}{1 - \frac{1}{8}} = \frac{2}{7}$$

$\therefore$ Probability that C wins

$$= 1 - \frac{4}{7} - \frac{2}{7} = \frac{1}{7}.$$

Ex. 7-88. *A and B toss a coin alternately on the understanding that the first to obtain head wins the toss. Show that their respective chances of winning are* $\frac{2}{3}$ *and* $\frac{1}{3}$.

Sol. Let (A) and (B) be the events that A and B get head in a toss respectively.

Then $$P(A) = P(B) = \frac{1}{2}$$

$\therefore$ $$P(\overline{A}) = P(\overline{B}) = 1 - \frac{1}{2} = \frac{1}{2}$$

Since A begins, he can win in following mutually exclusive ways :

$(A), (\overline{A}\,\overline{B}A), (\overline{A}\,\overline{B}\,\overline{A}\,\overline{B}A), ...$

Therefore, by theorem of total probability, probability that A wins

$= P(A) + P(\overline{A}\,\overline{B}A) + P(\overline{A}\,\overline{B}\,\overline{A}\,\overline{B}A) + ...$

$= P(A) + P(\overline{A})P(\overline{B})P(A) + \{P(\overline{A})P(\overline{B})\}^2 P(A) + ..$

$$= \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + ...$$

$$= \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}$$

$\therefore$ Probability that B wins

$$= 1 - \frac{2}{3} = \frac{1}{3}.$$

Ex. 7-89. *A coin is tossed three times. Find the chance that head and tail will show alternately.*

Sol. Let (H) and (T) denote the occurrence of head and tail in a toss respectively. There are two possibilities :

$$(H)(T)(H)$$

and $$(T)(H)(T)$$

Probability of either possibility

$$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Therefore, by total probability theorem probability of having head and tail alternately

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Ex. 7-90. *A and B are two independent witnesses (i.e., there is no collusion between them) in a case. The probability that A will speak the truth is x and the probability that B will speak the truth is y. A and B agree in a certain statement. Show that the probability that this statement is true is*

$$\frac{xy}{1 - x - y + 2xy}.$$

Sol. Let (A_1) and (A_2) b

statement is correct respectivel

Then P

and $P(A$

By compound probability t

$P(A$

$\therefore$ $P(A_2$

Ex. 7-91. *A coin is tossed*

consecutive heads is

Sol. Let p_n denote the pro

Two mutually exclusive ways o

(*i*) m consecutive heads o

(*ii*) Only $(m-1)$ consecut

consecutive heads, head must a

Probability of possibility (

and probability of possibility (

in last m trials).

By theorem of total proba

or

$\therefore$

nding that the first to obtain

ning are $\frac{2}{3}$ and $\frac{1}{3}$.

n a toss respectively.

ve ways :

at A wins

: that head and tail will show

il in a toss respectively. There

aving head and tail alternately

there is no collusion between
is x and the probability that B
. Show that the probability that

Sol. Let (A_1) and (A_2) be the events that A and B agree in a statement and their statement is correct respectively.

Then
$$P(A_1) = x.y+(1-x)(1-y)$$
$$= 1-x-y+2xy$$

and
$$P(A_1A_2) = xy$$

By compound probability theorem,
$$P(A_1A_2) = P(A_1)P(A_2/A_1)$$

$$\therefore \quad P(A_2/A_1) = \frac{P(A_1A_2)}{P(A_1)}$$

$$= \frac{xy}{1-x-y+2xy}$$

Ex. 7-91. *A coin is tossed (m+n) times (m>n). Show that the probability of getting m consecutive heads is*

$$\frac{n+2}{2^{m+1}}.$$

Sol. Let p_n denote the probability of getting m consecutive heads in $(m+n)$ tossings. Two mutually exclusive ways of having m consecutive heads in $(m+n)$ tossings are :

(*i*) m consecutive heads occur in $(m+n-1)$ tossings.

(*ii*) Only $(m-1)$ consecutive heads occur in $(m+n-1)$ tossings. In order to have m consecutive heads, head must appear in $(m+n)$th toss and there must be a tail in nth toss.

Probability of possibility $(i) = p_{n-1}$

and probability of possibility $(ii) =$ (Probability of tail in nth trial) × (Probability of heads in last m trials).

$$= \frac{1}{2}\cdot\left(\frac{1}{2}\cdot\frac{1}{2}\text{......}m \text{ times}\right)$$

$$= \frac{1}{2^{m+1}}$$

By theorem of total probability,

$$p_n = p_{n-1}+\frac{1}{2^{m+1}}$$

or
$$\frac{1}{2^{m+1}} = p_n - p_{n-1} = p_{n-1}-p_{n-2} = \ldots = p_1 - p_0$$

$$\therefore \quad p_n = p_0+\frac{n}{2^{m+1}}$$

p_0 = probability of getting m consecutive heads in $(m+0)$ tossings

$$= \frac{1}{2^m}$$

$$p_n = \frac{1}{2^m} + \frac{n}{2^{m+1}}$$

$$= \frac{n+2}{2^{m+1}}.$$

Ex. 7-92. (*a*) *What is the probability that two numbers chosen at random will be prime to each other ?*

(*b*) *Four positive integers are chosen at random. Find the probability of their having a common factor.*

Sol. (*a*) Let 'a' and 'b' be any two numbers and 'r' a prime number. When 'a' is divided by 'r' the possible remainders are

$$0, 1, 2, \ldots.(r-1)$$

Therefore, probability that 'a' is divisible by 'r'

i.e., probability of getting zero remainder $= \dfrac{1}{r}$.

Similarly probability that 'b' is divisible by 'r' $= \dfrac{1}{r}$

By compound probability theorem,

probability that 'r' divides both 'a' and 'b' $= \dfrac{1}{r^2}$.

Therefore, probability that 'a' and 'b' don't have a common factor 'r'

$$= \left(1 - \frac{1}{r^2}\right)$$

Therefore, probability that 'a' and 'b' are prime to each other

$$= \prod_r \left(1 - \frac{1}{r^2}\right) \qquad r = 2, 3, 5, 7, \ldots\ldots$$

$$= \frac{6}{\pi^2}$$

(*b*) Let a, b, c and d be four integers and r a prime number.

Then prob. that all the four numbers a, b, c, d are divisible by r

$$= \frac{1}{r^4}$$

Therefore prob. that a, b, c, d do not have r as a common factor

$$= \left(1 - \frac{1}{r^4}\right)$$

Therefore, prob. that four integers do not have any common factor

$$= \prod_r \left(1 - \frac{1}{r^4}\right) \qquad r = 2, 3, 5, 7 \ldots\ldots$$

Therefore, prob. that four

Ex. 7-93. *Cards are dealt*
Show that the probability that

If cards continue to be dea
exactly r cards are dealt in all

Sol. Since n cards are to b
from the remaining 48 cards le
$\therefore$ Probability for drawing

Evidently $(n+1)$th card n
$(52-n)$ cards.
Probability for having $(n$

By the theorem of compou
first ace

Out of r cards dealt before
$(n+1)$th draw.
Number of ways of dealir

$\therefore$ Probability of drawing

Evidently $(r+1)$th card r
$(52-r)$ cards.

$$= \frac{90}{\pi^4}$$

Therefore, prob. that four integers have a common factor

$$= 1 - \frac{90}{\pi^4}.$$

Ex. 7-93. *Cards are dealt one by one from a well shuffled pack until an ace appears. Show that the probability that exactly n cards are dealt before the first ace, is*

$$\frac{(51-n)(50-n)(49-n)}{13.51.50.49}$$

If cards continue to be dealt until a second ace appears, prove that the probability that exactly r cards are dealt in all before the second ace, is

$$\frac{r(51-r)(50-r)}{13.17.50.49}.$$

Sol. Since n cards are to be dealt before first ace appears, these n cards are to be dealt from the remaining 48 cards leaving aside 4 aces.

$\therefore$ Probability for drawing n cards (not containing any ace) from a pack

$$= \frac{{}^{48}C_n}{{}^{52}C_n}$$

Evidently $(n+1)$th card must be an ace to be drawn out of 4 aces present in remaining $(52-n)$ cards.

Probability for having $(n+1)$th card an ace

$$= \frac{{}^{4}C_1}{{}^{52-n}C_1}$$

By the theorem of compound probability, probability of dealing exactly n cards before first ace

$$= \frac{{}^{48}C_n}{{}^{52}C_n} \cdot \frac{{}^{4}C_1}{{}^{52-n}C_1}$$

$$= \frac{(51-n)(50-n)(49-n)}{13.51.50.49}$$

Out of r cards dealt before a second ace appears, one card is an ace which was drawn in $(n+1)$th draw.

Number of ways of dealing r cards containing one ace

$$= {}^{4}C_1 \; {}^{48}C_{r-1}$$

$\therefore$ Probability of drawing r cards containing one ace

$$= \frac{{}^{4}C_1 \; {}^{48}C_{r-1}}{{}^{52}C_r}$$

Evidently $(r+1)$th card must be an ace to be drawn out of 3 aces present in remaining $(52-r)$ cards.

Probability of having $(r+1)$th card a second ace

$$= \frac{^3C_1}{^{52-r}C_1}$$

Therefore, by compound probability theorem probability of dealing exactly r cards before second ace

$$= \frac{^4C_1 \, ^{48}C_{r-1}}{^{52}C_r} \cdot \frac{^3C_1}{^{52-r}C_1}$$

$$= \frac{r(51-r)(50-r)}{13.17.50.49}$$

Ex. 7-94. *Cards are dealt one-by-one from an ordinary pack (without replacements) until two aces have appeared. Find the most probable number of cards to be turned up.*

Sol. Let x be the total number of cards dealt until the second ace appears and $P(x)$ be its probability.

Evidently xth card must be an ace and among remaining $(x-1)$ cards there must be one ace.

Number of ways of drawing $(x-1)$ cards including one ace

$$= {^4C_1} \cdot {^{48}C_{x-2}}$$

$\therefore$ Prob. of drawing $(x-1)$ cards including one ace

$$= \frac{^4C_1 \, ^{48}C_{x-2}}{^{52}C_{x-1}}$$

Since xth card, which is to be an ace, is to be drawn from $52-(x-1)=53-x$ cards containing 3 aces, prob. that xth card is an ace

$$= \frac{3}{53-x}$$

Therefore, by compound probability theorem,

$$P(x) = \frac{^4C_1 \, ^{48}C_{x-2}}{^{52}C_{x-1}} \cdot \frac{3}{53-x}$$

$$= \frac{(x-1)(52-x)(51-x)}{13.17.50.49}.$$

Most probable number of cards is that value of x for which

$$P(x-1) < P(x) > P(x+1)$$

Consider $\qquad P(x-1) < P(x)$

i.e., $$\frac{(x-2)(53-x)(52-x)}{13.17.50.49} < \frac{(x-1)(52-x)(51-x)}{13.17.50.49}$$

i.e., $$x < \frac{55}{3}$$

Consider $\qquad P(x) > P(x+1)$

i.e.,
$$\frac{(x-1)(52-x)(51-x)}{13.17.50.49} > \frac{x(51-x)(50-x)}{13.17.50.49}$$

i.e.,
$$x > \frac{52}{3}$$

Therefore, most probable number x of cards is such that

$$\frac{52}{3} < x < \frac{55}{3}$$

Since x is to be an integer,

$$x = 18$$

$\therefore$ Most probable number of cards dealt

$$= 18.$$

Ex. 7-95. *Of three independent events the prob. that the first only should happen is* $\frac{1}{4}$ *the prob. that the second only should happen is* $\frac{1}{8}$ *and the prob. that the third only should happen is* $\frac{1}{12}$. *Obtain the unconditional probabilities of the three events.*

Sol. Let A_1, A_2 and A_3 be the events and p_1, p_2, p_3 be their unconditional probabilities.

Now
$$P(A_1\bar{A}_2\bar{A}_3) = \frac{1}{4},$$

i.e.,
$$P(A_1)P(\bar{A}_2)P(\bar{A}_3) = \frac{1}{4}$$

($\because A, B, C$ are independent)

i.e.,
$$p_1(1-p_2)(1-p_3) = \frac{1}{4} \qquad \ldots(1)$$

$$P(\bar{A}_1A_2\bar{A}_3) = \frac{1}{8}$$

i.e.,
$$(1-p_1)(p_2)(1-p_3) = \frac{1}{8} \qquad \ldots(2)$$

and
$$P(\bar{A}_1\bar{A}_2A_3) = \frac{1}{12}$$

i.e.,
$$(1-p_1)(1-p_2)p_3 = \frac{1}{12} \qquad \ldots(3)$$

From (1), (2) and (3)

$$p_1 = \frac{1}{2}, p_2 = \frac{1}{3} \text{ and } p_3 = \frac{1}{4}$$

Ex. 7-96. *An integer is chosen at random from the first two hundred integers, what is the prob. that the integers chosen is divisible by 6 or 8 ?*

Sol. Let A and B be the events that the number chosen is divisible by 6 and 8 respectively.

Since $6 \times 33 = 198$ and $8 \times 25 = 200$, there are 33 and 25 numbers upto 200 which are divisible by 6 and 8 respectively.

$$\therefore\ P(A) = \frac{33}{200}, \quad P(B) = \frac{25}{200}.$$

Now L.C.M. of 6 and 8 = 24.

The number which is divisible by 24 is divisible by 6 and 8 both.

$\therefore$ The number of numbers which are divisible by 6 and 8 both

$$= \text{greatest integer less than } \left\{\frac{200}{24}\right\}$$

$$= 8.$$

$$\therefore \qquad P(AB) = \frac{8}{200}$$

Now prob. that the integer chosen is divisible by 6 or 8

$$= P(A+B)$$

$$= P(A) + P(B) - P(AB)$$

$$= \frac{33}{200} + \frac{25}{200} - \frac{8}{200}$$

$$= \frac{50}{200} = \frac{1}{4}.$$

Ex. 7-97. *If n integers taken at random are multiplied together, show that (a) the chance that the last digit of the product is 1, 3, 7 or 9 is* $\left(\frac{2}{5}\right)^n$; *(b) the chance of its being 2, 4, 6 or 8 is* $\frac{4^n - 2^n}{5^n}$; *(c) of its being 5 is* $\frac{5^n - 4^n}{10^n}$; *and (d) of its being '0' is* $\frac{10^n - 8^n - 5^n + 4^n}{10^n}$.

Sol. (*a*) The last digit in the product will be 1, 3, 7 or 9 if each of the n integers end with either of these four digits. Now since there are in all 10 digits, prob. of selecting these four digits

$$= \frac{4}{10}$$

This is also the prob. for a integer to end with either 1, 3, 7 or 9.

$\therefore$ By compound prob. theorem, prob. that the last digit in the product of n integers is 1, 3, 7 or 9 = prob. that all n integers end with 1, 3, 7 or 9

$$= \left(\frac{4}{10}\right)^n = \left(\frac{2}{5}\right)^n.$$

(*b*) The last digit in the product will be 1, 2, 3, 4, 6, 7, 8 or 9 iff each of the n integers end with either of these eight digits.

$\therefore$ Prob. for an integer to end with either of these 8 digits.

$$= \frac{8}{10} = \frac{4}{5}.$$

$\therefore$ Prob. for the last digit in the product of n integers to be 1, 2, 3, 4, 6, 7, 8 or 9 $= \left(\frac{4}{5}\right)^n$.

Now the prob. for the last

{from (*a*)}.

$\therefore$ By total prob. theorem,

2, 4, 6, or 8 $= \frac{4^n - 2^n}{5^n}$.

(*c*) The last digit in the pı

with 1, 3, 7 and 9.

Now prob. that n integers

$\therefore$ Prob. that out of n integ

$= \left(\frac{5}{10}\right)^n$ − prob. that all intege

(*d*) Since total prob. is un

= 1−(prob. that last digit iı

= 1−(prob. that last digit iı

−(prob. that last digit in th

−(prob. that last digit in th

Ex. 7-98. *n letters to each c at random. (i) What is the pro What is the probability that ex*

Sol. Let U_n be the numbe

There are two mutually ex

(1) If any two letters occup

wrong in U_{n-2} ways. Since o

which all the letters can go wro

(2) If one letter occupies

$(n-1)$ ways, the remaining $(n$

Therefore, number of way

$\therefore$

or U_n

Change n to $n-1, n-2, \ldots$

Now the prob. for the last digit in the product of n integers to be 1, 3, 7 or 9 $= \left(\frac{2}{5}\right)^n$ {from (a)}.

$\therefore$ By total prob. theorem, prob. of having the last digit in the product of n integers to be 2, 4, 6, or 8 $= \dfrac{4^n - 2^n}{5^n}$.

(c) The last digit in the product will be 5 iff at least one integer end with 5 and others with 1, 3, 7 and 9.

Now prob. that n integers end with 1, 3, 5, 7 or 9.

$$= \left(\frac{5}{10}\right)^n.$$

$\therefore$ Prob. that out of n integers at least one integer ends with 5 and other with 1, 3, 7 or 9

$$= \left(\frac{5}{10}\right)^n - \text{prob. that all integers end with 1, 3, 7 or 9} = \left(\frac{5}{10}\right)^n - \left(\frac{4}{10}\right)^n = \frac{5^n - 4^n}{10^n}.$$

(d) Since total prob. is unity, prob. that the last digit in the product is '0'

= 1–(prob. that last digit in the product is 1, 2, 3, 4, 5, 6, 7, 8 or 9).

= 1–(prob. that last digit in the product is 1, 3, 7, or 9

–(prob. that last digit in the product is 2, 4, 6 or 8)

–(prob. that last digit in the product is 5)

$$= 1 - \left(\frac{4}{10}\right)^n - \left(\frac{4^n - 2^n}{5^n}\right) - \frac{5^n - 4^n}{10^n}$$

$$= \frac{1}{10^n}\{10^n - 8^n - 5^n + 4^n\}.$$

Ex. 7-98. *n letters to each of which corresponds an envelope are placed in the envelopes at random. (i) What is the probability that no letter is placed in the right envelope ? (ii) What is the probability that exactly r letters are placed in the right envelope ?*

Sol. Let U_n be the number of ways in which all the letters can go wrong.

There are two mutually exclusive possibilities :

(1) If any two letters occupy each other's position, the remaining $(n-2)$ letters can go wrong in U_{n-2} ways. Since one letter can go wrong in $(n-1)$ ways, number of ways in which all the letters can go wrong $= (n-1)\ U_{n-2}$.

(2) If one letter occupies another's envelope and not vice-versa which can happen in $(n-1)$ ways, the remaining $(n-1)$ letters can go wrong in U_{n-1} ways.

Therefore, number of ways in which all the letters can go wrong $= (n-1)U_{n-1}$.

$$\therefore \qquad U_n = (n-1)\{U_{n-1} + U_{n-2}\}$$

$$\text{or} \qquad U_n - nU_{n-1} = -\{U_{n-1} - (n-1)\,U_{n-2}\}$$

Change n to $n-1, n-2, \ldots..3$

$$U_{n-1}-(n-1)U_{n-2}=-\{U_{n-2}-(n-2)\,U_{n-3}\}$$

$$U_{n-2}-(n-2)\,U_{n-3}=-\{U_{n-3}-(n-3)\,U_{n-4}\}$$

....

$$U_3-3U_2=-\{U_2-2U_1\}$$

Multiplying

$$U_n-nU_{n-1}=(-1)^{n-2}\{U_2-2U_1\}$$

U_1 = number of ways in which one letter out of one can go wrong
$= 0$

and U_2 = number of ways in which two letters out of two can go wrong
$= 1.$

Therefore, $U_n-nU_{n-1}=(-1)^{n-2}=(-1)^n$

or
$$\frac{U_n}{n!}-\frac{U_{n-1}}{(n-1)!}=\frac{(-1)^n}{n!}$$

Change n to $n-1, n-2, \ldots, 2$

$$\frac{U_{n-1}}{(n-1)!}-\frac{U_{n-2}}{(n-2)!}=\frac{(-1)^{n-1}}{(n-1)!}$$

$$\frac{U_{n-2}}{(n-2)!}-\frac{U_{n-3}}{(n-3)!}=\frac{(-1)^{n-2}}{(n-2)!}$$

....

$$\frac{U_2}{2!}-\frac{U_1}{1!}=\frac{(-1)^2}{2!}$$

Adding $\dfrac{U_n}{n!}=\left\{\dfrac{1}{2!}-\dfrac{1}{3!}+\ldots\ldots+\dfrac{(-1)^{n-1}}{(n-1)!}+\dfrac{(-1)^n}{n!}\right\}$

Total number of ways of distributing n letters in n envelopes $= n!$

Therefore, reqd. prob. $=\dfrac{U_n}{n!}=\dfrac{1}{2!}-\dfrac{1}{3!}+\ldots+\dfrac{(-1)^n}{n!}$

(*ii*) r letters can be chosen out of n in ${}^nc_r$ ways and rest can go wrong in U_{n-r} ways.

Therefore, reqd. prob. $=\dfrac{{}^nc_rU_{n-r}}{n!}$

$$=\frac{{}^nc_r}{n!}\left[\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^{n-r}}{(n-r)!}\right](n-r)$$

$$=\frac{1}{r!}\left\{\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^{n-r}}{(n-r)!}\right\}$$

Ex. 7·99. *A player tosses*
two for every tail. He is to pla
attaining exactly n, show tha

and hence find the value of

Sol. Two mutually exclu
(1) when score is $(n-1)$
(2) when score is $(n-2$

The probabilities of these

and $\frac{1}{2}p_{n-2}$.

Therefore, by total prob

or

or $2(p_n$

or $(p_n$

Cha

$(p_{n}$

$(p_n$

.....

Multiplying p

Score '2' can be attaine
(1) tossing tail in the fi
(2) tossing heads in firs

Therefore,

Ex. 7-99. *A player tosses a coin and is to score one point for every head turned up and two for every tail. He is to play on until his score reaches or passes n. If* p_n *is the chance for attaining exactly n, show that*

$$p_n = \frac{1}{2}(p_{n-1} + p_{n-2})$$

and hence find the value of p_n.

Sol. Two mutually exclusive possibilities of attaining score exactly n are :

(1) when score is $(n-1)$, player tosses head.

(2) when score is $(n-2)$, player tosses tail.

The probabilities of these two possibilities, by compound probability theorem, are $\frac{1}{2}p_{n-1}$ and $\frac{1}{2}p_{n-2}$.

Therefore, by total probability theorem,

$$\begin{aligned} p_n &= \frac{1}{2}p_{n-1} + \frac{1}{2}p_{n-2} \\ &= \frac{1}{2}(p_{n-1} + p_{n-2}) \end{aligned}$$

or $$2p_n = p_{n-1} + p_{n-2}$$

or $$2(p_n - p_{n-1}) = -(p_{n-1} - p_{n-2})$$

or
$$\left.\begin{aligned} (p_n - p_{n-1}) &= \left(-\frac{1}{2}\right)(p_{n-1} - p_{n-2}) \\ &\text{Changing } n \text{ to } n-1, n-2, \ldots 3 \\ (p_{n-1} - p_{n-2}) &= \left(-\frac{1}{2}\right)(p_{n-2} - p_{n-3}) \\ (p_{n-2} - p_{n-3}) &= \left(-\frac{1}{2}\right)(p_{n-3} - p_{n-4}) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ (p_3 - p_2) &= \left(-\frac{1}{2}\right)(p_2 - p_1) \end{aligned}\right\} \quad \ldots(1)$$

Multiplying $$p_n - p_{n-1} = \left(-\frac{1}{2}\right)^{n-2}(p_2 - p_1)$$

Score '2' can be attained in following two mutually exclusive ways :
(1) tossing tail in the first trial.
(2) tossing heads in first two trials.

Therefore, $$p_2 = \frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}$$

$$= \frac{1}{2} + \frac{1}{4}$$

$$= \frac{3}{4}$$

Also $p_1 = \frac{1}{2}$ (probability of tossing head in first trial.)

Therefore, $p_n - p_{n-1} = \left(-\frac{1}{2}\right)^n$ ...(2)

Adding above equations (1)

$$p_n - p_2 = \left(-\frac{1}{2}\right)(p_{n-1} - p_1)$$

or
$$p_n - \frac{3}{4} = -\frac{1}{2} p_{n-1} + \frac{1}{4}$$

$$2p_n + p_{n-1} = 2 \quad ...(3)$$

Adding (2) and (3)

$$3p_n = 2 + \left(-\frac{1}{2}\right)^n$$

Therefore,
$$p_n = \frac{1}{3}\left\{2 + (-1)^n . \frac{1}{2^n}\right\}.$$

Ex. 7-100. *Each of the n urns contains a white and b black balls. One ball is transferred from the first urn into the second, then one ball from the latter into the third and so on. Finally one ball is taken from the last urn, what is the probability of its being white ?*

Sol. Let p_k be the probability of drawing a white ball from the kth urn.

There are two possibilities :

(1) A white ball is transferred from $(k-1)$th urn to kth urn.

(2) A black ball is transferred from $(k-1)$th urn to kth urn.

In (1) number of white balls in kth urn = $a+1$.

Therefore, conditional probability of drawing a white ball from kth urn when a white ball is transferred from $(k-1)$th urn to kth urn = $\frac{a+1}{a+b+1}$.

Probability of drawing a white ball from $(k-1)$th urn = p_{k-1}.

Therefore, by compound probability theorem, probability of drawing a white ball from kth urn. (If possibility (1) happens)

$$= \frac{a+1}{a+b+1} p_{k-1}.$$

Similarly probability of drawing a white ball from kth urn.

(If possibility (2) happens) = $\frac{a}{a+b+1}(1 - p_{k-1})$

Therefore, by theorem of total

$-\frac{1}{2} p_{n-1} + \frac{1}{4}$

Put

But p_1 = probability of drawi

Therefore,

Similarly,

In general,

Ex. 7-101. *In a lottery m-ticke and returned before the next drawi each of the numbers 1, 2....n will a*

$p_k = 1 - {}^n c_1 \left(1 - \right.$

Sol. Let $(A_1), (A_2), \ldots\ldots (A_n$

appear at least once in k drawing. numbers, 1, 2,.....n respectively do

By additive law,

$P(\overline{A}_1 + \overline{A}_2 + \ldots\ldots$

Probability of appearance of i

Therefore, probability of non-

Therefore, by theorem of total probability

$$-\frac{1}{2}p_{n-1}+\frac{1}{4}$$

$$p_k = \frac{a+1}{a+b+1}p_{k-1}+\frac{a}{a+b+1}(1-p_{k-1})$$

$$= \frac{1}{a+b+1}p_{k-1}+\frac{a}{a+b+1}$$

Put $k = 2$

$$p_2 = \frac{1}{a+b+1}p_1+\frac{a}{a+b+1}$$

But $p_1 =$ probability of drawing a white ball from first urn.

$$= \frac{a}{a+b}$$

Therefore, $$p_2 = \frac{1}{a+b+1}\cdot\frac{a}{a+b}+\frac{a}{a+b+1}$$

$$= \frac{a}{a+b}$$

Similarly, $p_3 = \dfrac{a}{a+b}$ and so on.

In general,

$$p_n = \frac{a}{a+b}.$$

Ex. 7-101. *In a lottery m-tickets are drawn at a time out of the total number of n tickets and returned before the next drawing is made. Show that the probability that in k drawings each of the numbers 1, 2....n will appear at least once is*

$$p_k = 1-{}^nc_1\left(1-\frac{m}{n}\right)^k+{}^nc_2\left(1-\frac{m}{n}\right)^k\left(1-\frac{m}{n-1}\right)^k \ldots\ldots$$

Sol. Let $(A_1), (A_2), \ldots\ldots\ldots (A_n)$ denote the events that the numbers, 1, 2,....n respectively appear at least once in k drawing. Then $(\bar{A}_1), (\bar{A}_2), \ldots\ldots (\bar{A}_n)$ denote the events that the numbers, 1, 2,.....n respectively do not appear in k drawings.

By additive law,

$$P(\bar{A}_1+\bar{A}_2+\ldots\ldots\bar{A}_n) = \sum_{i=1}^{n}P(A_i)-\sum_{\substack{i,j=1\\ i<j}}^{n}P(\bar{A}_i\bar{A}_j)+\ldots\ldots$$

Probability of appearance of ith number in one draw

$$= \frac{m}{n}$$

Therefore, probability of non-appearance of ith number in one draw

$$= \left(1-\frac{m}{n}\right)$$

Since tickets are replaced after each draw, draws are independent and hence by compound probability theorem

$p(\overline{A}_i) =$ Probability of non-appearance of ith number in k drawings

$$= \left(1 - \frac{m}{n}\right)^k$$

Therefore, $\sum_{i=1}^{n} P(\overline{A}_i) = {}^n c_1 \left(1 - \frac{m}{n}\right)^k.$

Probability of non-appearance of any two specified numbers in one draw

$$= \frac{{}^{n-2}c_m}{{}^n c_m}$$

$$= \frac{(n-m)(n-m-1)}{n(n-1)}$$

$$= \left(1 - \frac{m}{n}\right)\left(1 - \frac{m}{n-1}\right)$$

Therefore, $P(\overline{A}_i A_j) =$ Probability of non-appearance of ith and jth numbers in k drawings

$$= \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k$$

Therefore, $$\sum_{\substack{i,j=1\\ i<j}}^{n} P(\overline{A}_i \overline{A}_j) = {}^n c_2 \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k$$

Similarly,

$$\sum_{\substack{i,j,k=1\\ i<j<k}}^{n} P(\overline{A}_i \overline{A}_j \overline{A}_k) = {}^n c_3 \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k \left(1 - \frac{m}{n-2}\right)^k$$

and so on.

Therefore, $P(\overline{A}_1 + \overline{A}_2 + \ldots + \overline{A}_n)$

$$= {}^n c_1 \left(1 - \frac{m}{n}\right)^k - {}^n c_2 \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k$$

$$+ {}^n c_3 \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k \left(1 - \frac{m}{n-2}\right)^k - \ldots\ldots$$

Therefore, $P(A_1 A_2 \ldots A_n) = 1 - P(\overline{A}_1 + \ldots + \overline{A}_n)$

$$= 1 - {}^n c_1 \left(1 - \frac{m}{n}\right)^k + {}^n c_2 \left(1 - \frac{m}{n}\right)^k \left(1 - \frac{m}{n-1}\right)^k$$

Which is the required prc

Ex. 7-102. *We have k vari objects. These objects are dra that the probability p_n that n all varieties is given by*

$k^{n-1} p_n = (k -$

Sol. In the first draw ther drawing $(k-1)$ varieties in (probability of at least one out o

Let $(B_1), (B_2), \ldots (B_{k-1})$ are absent respectively. Evide

Therefore by additive law

$P(B_1 + B_2 + \ldots\ldots + B_{k-1}) =$

If (B_1) happens, out of drawn in first draw) are missing above) in a draw

Therefore, $P(B_1) =$ Proba draws

Similarly $P(B_2) = P(B_3)$

If $(B_1 B_2)$ happens out o drawn in first draw) are missir

Therefore as above,

$P($

and so on.

$\therefore$ $P(B_1 + B_2 \ldots\ldots + B_{k-1})$

and so on.

ndent and hence by compound

k drawings

pers in one draw

and jth numbers in k drawings

$$\left(1-\frac{m}{n-2}\right)^k$$

$$-\frac{m}{n}\Big)^k\left(1-\frac{m}{n-1}\right)^k$$

$$\frac{n}{-1}\Big)^k\left(1-\frac{m}{n-2}\right)^k-\ldots\ldots$$

$$2\left(1-\frac{m}{n}\right)^k\left(1-\frac{m}{n-1}\right)^k$$

$$-{}^nc_3\left(1-\frac{m}{n}\right)^k\left(1-\frac{m}{n-1}\right)^k\left(1-\frac{m}{n-2}\right)^k+\ldots..$$

Which is the required probability.

Ex. 7-102. *We have k varieties of objects each variety consisting of the same number of objects. These objects are drawn one at a time and replaced before the next drawing. Show that the probability* p_n *that n and no less drawings will be required to produce objects of all varieties is given by*

$$k^{n-1}p_n=(k-1)^{n-1}-{}^{k-1}c_1(k-2)^{n-1}+{}^{k-1}c_2(k-3)^{n-1}\ldots\ldots$$

Sol. In the first draw there will be necessarily one variety, therefore, the probability of drawing $(k-1)$ varieties in $(n-1)$ drawings is to be obtained. To proceed with firstly the probability of at least one out of $(k-1)$ varieties missing in $(n-1)$ drawings will be obtained.

Let $(B_1),(B_2),\ldots.(B_{k-1})$ denote the events that first, second.......... $(k-1)$th varieties are absent respectively. Evidently events $(B_1),(B_2)\ldots(B_{k-1})$ are non-mutually exclusive.

Therefore by additive law,

$$P(B_1+B_2+\ldots\ldots+B_{k-1})=\sum_{i=1}^{k-1}P(B_i)-\sum_{\substack{i,j=1\\i<j}}^{k-1}P(B_iB_j)+\ldots\ldots$$

If (B_1) happens, out of k varieties two varieties (first variety and variety which was drawn in first draw) are missing. Probability of not drawing any of the two varieties (mentioned above) in a draw

$$=\left(\frac{k-2}{k}\right)$$

Therefore, $P(B_1)=$ Probability that two varieties (mentioned above) are absent in $(n-1)$ draws

$$=\left(\frac{k-2}{k}\right)^{n-1}$$

Similarly $P(B_2)=P(B_3)=\ldots=P(B_{k-1})=\left(\frac{k-2}{k}\right)^{n-1}$

If (B_1B_2) happens out of k varieties three varieties (first, second and one which was drawn in first draw) are missing.

Therefore as above,

$$P(B_1B_2)\;=\;P(B_1B_3)=\ldots=\left(\frac{k-3}{k}\right)^{n-1}$$

and so on.

$$\therefore\;P(B_1+B_2\ldots\ldots+B_{k-1})={}^{k-1}c_1\left(\frac{k-2}{k}\right)^{n-1}-{}^{k-1}c_2\left(\frac{k-3}{k}\right)^{n-1}+\ldots.$$

and so on.

Also probability that in $(n-1)$ draws the variety that has been drawn in first draw is absent $=\left(1-\dfrac{1}{k}\right)^{n-1}$

$$\therefore \text{ Reqd. prob.} = \left(1-\frac{1}{k}\right)^{n-1} - P(B_1+B_2+\ldots+B_{k-1})$$

$$\therefore\ p_n = \left(\frac{k-1}{k}\right)^{n-1} - {}^{k-1}c_1\left(\frac{k-2}{k}\right)^{n-1} + {}^{k-1}c_2\left(\frac{k-3}{k}\right)^{n-1} \ldots.$$

or $k^{n-1}p_n = (k-1)^{n-1} - {}^{k-1}c_1(k-2)^{n-1} + {}^{k-1}c_2(k-3)^{n-1}\ldots.$

Ex. 7-103. *Two similar decks of n different cards are put into random order and matched against each other. If a card occupies the same position in both the decks we speak of match (coincidence). Find the probability of at least one match.*

Sol. Assume that cards of one deck be in natural order. Let (A_i) be the event that a match occurs at the *i*th place.

By additive law,

$$P(A_1+A_2+\ldots+A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j=1}^{n} P(A_iA_j) + \ldots + (-1)^{n-1}P(A_1A_2\ldots A_n)$$

If event (A_i) happens, a match occurs at the *i*th place, *i.e.*, in second deck *i*th numbered card is at the *i*th place while the remaining $(n-1)$ cards may be in any order.

Number of ways of distributing $(n-1)$ cards on $(n-1)$ places

$$= (n-1)$$

and total of number ways of distributing n cards on n places

$$= n!$$

$$\therefore \qquad P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

Similarly,
$$P(A_iA_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

and so on.

$$\therefore \qquad P(A_1+A_2+\ldots\ldots\ldots\ldots+A_n) = {}^nc_1\cdot\frac{1}{n} - {}^nc_2\cdot\frac{1}{n(n-1)}$$

$$+\,{}^nc_3\,\frac{1}{n(n-1)(n-2)} + \ldots. + (-1)^{n-1}\frac{1}{n!}$$

$$= 1-\frac{1}{2!}+\frac{1}{3!}+\ldots+\frac{(-1)^{n-1}}{n!}.$$

Ex. 7-104. *What is the probability that at least one of the players in a bridge game will get a complete suit of cards ?*

Sol. Let $(A_1),(A_2),(A_3)$ and (A_4) denote the events that four players respectively get a complete suit of cards

been drawn in first draw is

random order and matched
'he decks we speak of match

et (A_i) be the event that a

$)^{n-1} P(A_1 A_2 ... A_n)$

n second deck ith numbered
: in any order.
aces

:s

$$\frac{1}{n(n-1)}$$

'ayers in a bridge game will

our players respectively get

$P(A_1) =$ Probability that player gets a complete suit of cards

$$= \frac{{}^4c_1}{{}^{52}c_{13}}$$

as there are four suits and player is to get one.

Similarly $P(A_2) = P(A_3) = P(A_4) = \frac{{}^4c_1}{{}^{52}c_{13}}$

$P(A_i A_j) =$ Probability that ith and jth player get complete suit of cards

$$= \frac{{}^4c_1}{{}^{52}c_{13}} \frac{{}^3c_1}{{}^{39}c_{13}}$$

$$i = 1, 2, 3, 4$$
$$j = 1, 2, 3, 4 \text{ and } i \neq j$$

Similarly $\quad P(A_i A_j A_k) = \frac{{}^4c_1}{{}^{52}c_{13}} \cdot \frac{{}^3c_1}{{}^{39}c_{13}} \cdot \frac{{}^2c_1}{{}^{26}c_{13}} \; i, j, k = 1, 2, 3, 4 \;\&\; i \neq j \neq k$

and $\quad P(A_1 A_2 A_3 A_4) = \frac{{}^4c_1}{{}^{52}c_{13}} \cdot \frac{{}^3c_1}{{}^{39}c_{13}} \cdot \frac{{}^2c_1}{{}^{26}c_{13}} \cdot \frac{1}{{}^{13}c_{13}}$

$$P(A_1 + A_2 + A_3 + A_4) = \sum_{i=1}^{4} P(A_i) - \sum_{i<j=1}^{4} P(A_i A_i)$$

$$+ \sum_{i<j<k=1}^{4} P(A_i A_j A_k) - P(A_1 A_2 A_3 A_4)$$

$$= {}^4c_1 \cdot \frac{{}^4c_1}{{}^{52}c_{13}} - {}^4c_2 \cdot \frac{{}^4c_1 . {}^3c_1}{{}^{52}c_{13} . {}^{39}c_{13}} + {}^4c_3 \frac{{}^4c_1 . {}^3c_1 . {}^2c_1}{{}^{52}c_{13} . {}^{39}c_{13} . {}^{26}c_{13}}$$

$$- \frac{{}^4c_1 . {}^3c_1 . {}^2c_1 . 1}{{}^{52}c_{13} . {}^{39}c_{13} . {}^{26}c_{13} \; {}^{13}c_{13}}$$

$$= 16.\frac{13!.39!}{(52)!} - \frac{6.4.3.(13!)^2}{(52)!} 26! + \frac{4.4.3.2}{(52)!}(13!)^4 - 24\frac{(13!)^4}{(52)!}$$

$$= \frac{16.13!39! - 72(13!)^2 26! + 72(13!)^4}{52!}$$

Which is the required probability.

Ex. 7-105. *Show that*

$$P(AB) \leq P(A) \leq P(A+B) \leq P(A) + P(B).$$

Sol. By compound prob. theorem,

$$P(AB) = P(A)P(B/A)$$

Since $\quad P(B/A) \leq 1, P(AB) \leq P(A) \quad$ ...(1)

Since $A\overline{B}, \overline{A}B, AB$ are mutually exclusive forms in which an event $(A+B)$ can happen, by total prob. theorem,

$$P(A+B) = P(A\overline{B}) + P(\overline{A}B) + P(AB)$$

Similarly $\quad P(A) = P(A\overline{B}) + P(AB)$

$\therefore \quad P(A+B) = P(A) + P(\overline{A}B)$

Since $\quad P(\overline{A}B) \geq 0, P(A) \leq P(A+B) \qquad \ldots(2)$

Also $\quad P(A+B) = P(A) + P(B) - P(AB)$

Since $\quad P(AB) \geq 0, P(A+B) \leq P(A) + P(B) \qquad \ldots(3)$

Result follows from (1), (2) and (3).

Ex. 7-106. State and prove Bayes theorem.

Sol. Statement. If an event E_1 can only occur in combination with one of the mutually exclusive events $E_1, E_2 \ldots\ldots\ldots E_n$, then

$$P(E_k / E) = \frac{P(E_k)P(E/E_k)}{\sum_{i=1}^{n} P(E_i)P(E/E_i)} \qquad k = 1, 2, \ldots\ldots n$$

Proof. Since the events E can occur only with the events $E_1, E_2, \ldots\ldots\ldots E_n$, the possible forms in which E can occur are

$$EE_1, EE_2, \ldots\ldots\ldots EE_n.$$

These forms are mutually exclusive as the events E are mutually exclusive.

$\therefore$ By total prob. theorem,

$$P(E) = P(EE_1) + (EE_2) + \ldots. P(EE_n)$$

$$= \sum_{i=1}^{n} P(EE_i) = \sum_{i=1}^{n} P(E_i)P(E/E_i)$$

(using compound prob. theorem)

Now by compound prob. theorem

$$P(EE_k) = P(E)P(E_k/E) = P(E_k)P(E/E_k)$$

$$\therefore \quad P(E_k/E) = \frac{P(E_k)P(E/E_k)}{P(E)}$$

$$= \frac{P(E_k)P(E/E_k)}{\sum_{i=1}^{n} P(E_i)P(E/E_i)}$$

Note. The probabilities $P(E_k)$ and $P(E_k/E)$ are known as '*priori*' and '*posteriori*' probabilities.

Thus, Baye's theorem can be stated as :

'*If an event E can occur only in combination with the mutually exclusive events $E_1, E_2, \ldots E_n$ and if*

(i) the priori probabilities $P(E_1), P(E_2), \ldots P(E_n)$ *corresponding to the total absence of knowledge regarding the occurrence of E and (ii) the conditional probabilities*

$$P(E/E_1), P(E/E_2)\ldots\ldots, P(E/E_n)$$

are given, the posteriori probabilities

$$P(E_1/E), P(E_2/E)\ldots\ldots, P(E_n/E)$$

are given by

$$P(E_k/E) = \frac{P(E_k)P(E/E_k)}{\sum_{i=1}^{n} P(E_i)P(E/E_i)} \qquad k = 1, 2, \ldots n$$

Ex. 7-107. *A bridge player knows that his opponents have exactly five hearts between the two of them. Each opponent has thirteen cards. What is the probability that there is a three-two split on the hearts (i.e., one player has three hearts and the other two) ?*

Sol. Let E_1: the event that two opponents have exactly five hearts between them.

E_2: the event that one player has three hearts and the other two.

then $P(E_2/E_1)$ is to be obtained.

$$P(E_2/E_1) = \frac{P(E_1E_2)}{P(E_1)}$$

Now
$$P(E_1) = \frac{\left({}^{13}C_5\ {}^{39}C_{21}\right)}{{}^{52}C_{26}}$$

and
$$P(E_1E_2) = \frac{\left({}^{13}C_3 \cdot {}^{39}C_{10}\right)\left({}^{10}C_2 \cdot {}^{29}C_{11}\right) + \left({}^{13}C_2 \cdot {}^{39}C_{11}\right)\left({}^{11}C_3 \cdot {}^{28}C_{10}\right)}{{}^{52}C_{13} \cdot {}^{39}C_{13}}$$

$$= \frac{2 \cdot \left({}^{13}C_3 \cdot {}^{39}C_{10}\right)\left({}^{10}C_2 \cdot {}^{29}C_{11}\right)}{{}^{52}C_{13} \cdot {}^{39}C_{13}}$$

$\therefore$
$$P(E_2/E_1) = \frac{2 \cdot {}^{13}C_3 \cdot {}^{39}C_{10} \cdot {}^{10}C_2 \cdot {}^{29}C_{11} \cdot {}^{52}C_{26}}{{}^{52}C_{13} \cdot {}^{39}C_{13} \cdot {}^{13}C_5 \cdot {}^{39}C_{21}}$$

$$= \frac{78}{115}.$$

Ex. 7-108. *Urn A contains two white and two black balls; Urn B contains three white and two black balls. One ball is transferred from A to B; one ball is then drawn from B and turns out to be white. What is the probability that the transferred ball was white ?*

Sol. Let E_1: event that transferred ball was white

E_2: event that transferred ball was black

E_3: a white ball is drawn from B

$P(E_1/E_3)$ is to be obtained.

Now $P(E_1)=\frac{2}{4}=\frac{1}{2}$, $P(E_2)=\frac{2}{4}=\frac{1}{2}$

$$P(E_3/E_1)=\frac{4}{6}=\frac{2}{3},\quad P(E_3/E_2)=\frac{3}{6}=\frac{1}{2}$$

By Bay's theorem,

$$P(E_1/E_3) = \frac{P(E_3/E_1)\,P(E_1)}{P(E_3/E_1)P(E_1)+P(E_3/E_2)P(E_2)}$$

$$= \frac{\frac{2}{3}\times\frac{1}{2}}{\frac{2}{3}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}}=\frac{4}{7}.$$

Ex. 109. *Die A has four red and two blue faces and die B has two red and four blue faces. The following game is played : First a coin is tossed once. If it falls heads, the game continues by repeatedly throwing die A; if it falls tails die B is repeatedly tossed.*

(*a*) *Show that the probability of red in any throw is* $\frac{1}{2}$.

(*b*) *If the first two throws of the die resulted in red, what is the probability of red at the third throw ?*

(*c*) *If the red turns up at the first n throws, what is the probability that die A is being used ?*

Sol. (*a*) At any throws, there are two possibilities :

(*i*) Coin shows head and die A is used.

Here prob. of red $=\frac{1}{2}\times\frac{4}{6}=\frac{1}{3}$.

(*ii*) Coin shows tail and die B is used.

Here prob. of red $=\frac{1}{2}\times\frac{2}{6}=\frac{1}{6}$.

$\therefore$ Prob. of red at any throw $=\frac{1}{3}+\frac{1}{6}=\frac{1}{2}$

(*b*) Let E_1: first two throws of the die resulted in red

E_2: red at the third throw.

To find $P(E_1)$,

if die A is used, prob. of red $=\left(\frac{2}{3}\right)^2$

if die B is used, prob. of red $=\left(\frac{1}{3}\right)^2$

$$\therefore\quad P(E_1) = \frac{1}{2}\times\left(\frac{2}{3}\right)^2+\frac{1}{2}\times\left(\frac{1}{3}\right)^2=\frac{5}{18}$$

To find $P(E_1E_2)$

if die A is used,

if die B is used,

(*c*) Let E: red turns up a

Ex. 7-110. *In a bolt fac per cent of the total. Out of drawn from the produce ar manufactured by A, B and C*

Sol. Let E be the event t is being produced by A, B, C

Then $P(E_1)=0\cdot25$, $P($

and $P(E/E_3)=0\cdot02$.

It is required to find $P($

By Baye's theorem

$P(E_1/E) =$

To find $P(E_1 E_2)$

$$\text{if die } A \text{ is used, prob. of red } = \left(\frac{2}{3}\right)^3$$

$$\text{if die } B \text{ is used, prob. of red } = \left(\frac{1}{3}\right)^3$$

$$\therefore \qquad P(E_1 E_2) = \frac{1}{2}\left\{\left(\frac{2}{3}\right)^3 + \left(\frac{1}{3}\right)^3\right\} = \frac{9}{54}$$

$$\therefore \qquad P(E_2 / E_1) = \frac{P(E_1 E_2)}{P(E_1)} = \frac{9/54}{5/18} = \frac{3}{5}$$

(c) Let E: red turns up at the first n throws

$$P(E / A) = \left(\frac{2}{3}\right)^n$$

$$P(E / B) = \left(\frac{1}{3}\right)^n$$

$$P(A / E) = \frac{P(E / A)\, P(A)}{P(E / A)\, P(A) + P(E / B)\, E(B)}$$

$$= \frac{\left(\frac{2}{3}\right)^n \cdot \frac{1}{2}}{\left(\frac{2}{3}\right)^n \cdot \frac{1}{2} + \left(\frac{1}{3}\right)^n \cdot \frac{1}{2}} = \frac{1}{1 + 2^{-n}}$$

Ex. 7-110. *In a bolt factory machines A, B, C manufacture respectively 25, 35 and 40 per cent of the total. Out of their output 5, 4 and 2 per cent are defective bolts. A bolt is drawn from the produce and is found defective. What are the probabilities that it was manufactured by A, B and C ?*

Sol. Let E be the event that the bolt is defective and E_1, E_2, E_3 the events that the bolt is being produced by A, B, C respectively.

Then $P(E_1) = 0{\cdot}25$, $P(E_2) = 0{\cdot}35$, $P(E_3) = 0{\cdot}40$, $P(E / E_1) = 0{\cdot}05$, $P(E / E_2) = 0{\cdot}04$ and $P(E / E_3) = 0{\cdot}02$.

It is required to find $P(E_1 / E)$, $P(E_2 / E)$ *and* $P(E_3 / E)$

By Baye's theorem

$$P(E_1 / E) = \frac{P(E_1)P(E / E_1)}{P(E_1)P(E / E_1) + P(E_2)P(E / E_2) + P(E_3)P(E / E_2)}$$

$$= \frac{(0{\cdot}25)\,(0{\cdot}05)}{(0{\cdot}25)\,(0{\cdot}05) + (0{\cdot}35)\,(0{\cdot}04) + (0{\cdot}4)\,(0{\cdot}02)}$$

$$= \frac{125}{345}$$

Similarly, $$P(E_2 / E) = \frac{140}{345}$$

and $$P(E_3 / E) = \frac{80}{345}.$$

Ex. 7-111. *Prove that*

$P(A_1 + A_2 + ...A_n) \le P(A_1) + P(A_2)....+P(A_n)$ (Bool's inequality)

Proof. We have

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$$

$$\therefore \quad P(A_1 + A_2) \le P(A_1) + P(A_2)$$

$$\therefore \quad P(A_1 + A_2 + A_3) = P(A_1 + \overline{A_1 + A_2})$$
$$\le P(A_1) + P(A_2 + A_3)$$
$$\le P(A_1) + P(A_2) + P(A_3)$$

Let $P(A_1 + A_2 + ...+A_m) \le P(A_1) + P(A_2) + ...+P(A_m)$

Then $P(A_1 + A_2 + ...+A_{m+1}) = P(A_1 + \overline{A_2 + ...+A_{m+1}})$

$$\le P(A_1) + P(A_2 + A_3 + ...+A_{m+1})$$
$$\le P(A_1) + \{P(A_2) + P(A_3) + ...+P(A_{m+1})\}$$
$$= P(A_1) + P(A_2) + ...+P(A_{m+1})$$

$\therefore$ By induction result follows :

Ex. 7-112. *For n events* $A_1, A_2 ....A_n$ *show that* $P(A_1 A_2 ... A_n) \ge P(A_1) + P(A_2) + ... + P(A_n) - (n-1)$

Sol. For two events A_1 and A_2 we have

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$$

Also $P(A_1 + A_2) \le 1$

$$\therefore \quad P(A_1) + P(A_2) - P(A_1 A_2) \le 1$$

$$\Rightarrow \quad P(A_1 A_2) \ge P(A_1) + P(A_2) - 1 \quad ...(1)$$

Result is true for $n = 2$.
Let result is true for $n = m$
Then

$$P(A_1 A_2 ... A_m) \ge P(A_1) + ... + P(A_m) - (m-1) \quad ...(2)$$

Now $P(A_1 A_2 ... A_m A_{m+1}) = P(A_1 A_2 ... A_m) A_{m+1})$

$$\ge P(A_1 A_2 ... A_m) + P(A_{m+1}) - 1 \quad \text{(by (1))}$$

$$\ge \{P(A_1) + P(A_2) + ... + P(A_m) - (m-1)\} + P(A_{m+1}) - 1 \quad \text{(using 2)}$$

$$= P(A_1) + ...+P(A_{m+1}) - (\overline{m+1} - 1)$$

$\therefore$ If result is true for $n =$
$\therefore$ Result is true for any p

Ex. 7-113. *If* $A_1 A_2 ......A_n$

$P(A_1$

Sol.

$P(A_1 + ...+A$

Ex. 7-114. *In a toss of tw the first coin (ii) two heads gi*

Sol. Let A_1, A_2 be the ev

Then

(*i*) $P(A_1 A_2$

(*ii*) $P(A_1 A_2 / A_1$

Ex. 7-115. *There are five 'i' has 'i' defective balls and*

(*i*) *An urn is selected at r drawn is defective.*

(*ii*) *A ball is drawn from on that it came from urn 5.*

(Bool's inequality)

$A_2)$

$)$

$A_m)$

$_1)$

$....+A_{m+1})$

$)+.......+P(A_{m+1})\}$

$P(A_{m+1})$

$_2...A_n) \geq P(A_1) + P(A_2)+...$

$A_2)$

...(1)

$-1)$ ...(2)

$_1)-$ (by (1))

(using 2)

$-(\overline{m+1}-1)$

$\therefore$ If result is true for $n = m$ then it is also true for $n = m+1$

$\therefore$ Result is true for any positive integral value of n.

Ex. 7-113. *If* $A_1 A_2, ...A_n, B$ *are events and* $P(B) > 0$ *show that*

$$P(A_1 + A_2 + ...+A_n / B) \leq \sum_{i=1}^{n} P(A_i / B)$$

Sol.

$$P(A_1 + ...+A_n / B) = \frac{P\{(A_1 + ...+A_n)B\}}{P(B)}$$

$$= \frac{P(A_1 B + ...+A_n B)}{P(B)}$$

$$\leq \sum_i \frac{P(A_i B)}{P(B)}$$

$$= \sum_i P(A_i / B)$$

Ex. 7-114. *In a toss of two coins, find the probability of (i) two heads given a head on the first coin (ii) two heads given at least one head.*

Sol. Let A_1, A_2 be the events of head on first and second coin respectively.

Then $$P(A_1) = P(A_2) = \frac{1}{2}$$

(*i*) $$P(A_1 A_2 / A_1) = \frac{P(A_1 A_2 A_1)}{P(A_1)} = \frac{P(A_1 A_2)}{P(A_1)}$$

$$= \frac{P(A_1)\, P(A_2)}{P(A_1)} = P(A_2) = \frac{1}{2}$$

(*ii*) $$P(A_1 A_2 / A_1 \cup A_2) = \frac{P(A_1 A_2 A_1 \cup A_2)}{P(A_1 \cup A_2)}$$

$$= \frac{P(A_1 A_2)}{P(A_1) + P(A_2) - P(A_1 A_2)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} + \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2}} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Ex. 7-115. *There are five urns numbered from 1 to 5 and each containing 10 balls. Urn 'i' has 'i' defective balls and 10-i non-defective balls (i=1, 2.....5) :*

(*i*) *An urn is selected at random and a ball is drawn. Find the probability that the ball drawn is defective.*

(*ii*) *A ball is drawn from one of the urns and it is found to be defective. Find the probability that it came from urn 5.*

Sol. Let A = Event that a defective ball is selected.

B_i = Urn i is selected.

Then $\quad P(B_i) = \dfrac{1}{5} \qquad i = 1, 2, \ldots\ldots 5$

$$P(A / B_i) = \frac{i}{10}, \text{ (as } B_i \text{ contains } i \text{ defective balls)}$$

(*i*) $$P(A) = \sum_{i=1}^{5} P(A / B_i)\, P(B_i)$$

$$= \frac{1}{5}\left\{\sum_{i=1}^{5} \frac{i}{10}\right\} = \frac{1}{50} \cdot \frac{5 \cdot 6}{2} = \frac{3}{10}$$

(*ii*) $$P(B_5 / A) = \frac{P(A / B_5)\, P(B_5)}{\sum\limits_{i=1}^{5} P(A / B_i)\, P(B_i)} = \frac{\frac{5}{10} \cdot \frac{1}{5}}{3/10} = \frac{1}{3}.$$

Ex. 7-116. *An urn contains 3 black and 7 white balls. At each trial a ball is selected at random, its colour is noted and it is replaced along with two additional balls of the same colour. Find the probability that a black ball is selected in each of the first three trials.*

Sol. Let B_i = Event that a black ball is selected in ith trial.

Then $P(B_1) = \dfrac{3}{10}$

and given that B_1 has happened, no. of black balls before second trial = $3 + 2 = 5$ and total no. of balls = 12.

$$\therefore \quad P(B_2 / B_1) = \frac{5}{12}$$

and given that $B_1 B_2$ has happened, no. of black balls before 3rd trial = 5 + 2 = 7 and total no. of balls = 14.

$$\therefore \quad P(B_3 / B_1 B_2) = \frac{7}{14}$$

$$\therefore \quad P(B_1 B_2 B_3) = P(B_1 B_2)\, P(B_3 / B_1 B_2)$$

$$= P(B_1)\, P(B_2 / B_1)\, P(B_3 / B_1 B_2)$$

$$= \frac{3}{10} \cdot \frac{5}{12} \cdot \frac{7}{14} = \frac{1}{16}.$$

Ex. 7-117. *An urn contains M balls numbered 1 to M. Where the first K balls are defective and the remaining M–K balls are non-defective, n balls are drawn from the urn one-by-one. Find the probability that the sample of n balls contains exactly k defective balls when sampling is (i) with replacement, (ii) without replacement.*

Sol. (*i*) *With replacement*

Let z_j = no. of the ball dra

Then sample space S is

and (

Let A_k = The event that t

Then A_k = Subset of S for

the remaining $(n - k)$ z_j's are s.t

Now in n-places, k–z_j's (1 ≤

chosen in K^k ways and remaini

∴ O(.

∴ P(

(*ii*) *Without replacement*

Here C

and O(.

∴ P(

Ex. 7-118. *An urn contains*

of size n is drawn. Find the prob

contains k black balls.

Sol. Two possibilities are th

(*i*) sampling with replacem

(*ii*) sampling without replac

Let A_k = Event that sampl

B_j = Event that jth bal

(*i*) *With replacement*

Here P(

P(A_k /

P(

i defective balls)

$\frac{6}{} = \frac{3}{10}$

$\frac{\frac{5}{10} \cdot \frac{1}{5}}{3/10} = \frac{1}{3}$.

each trial a ball is selected at
ɔ additional balls of the same
ıch of the first three trials.
ial.

ond trial $= 3+2=5$ and total

3rd trial $= 5+2=7$ and total

)

$_3 / B_1 B_2)$

. Where the first K balls are
balls are drawn from the urn
ıtains exactly k defective balls
ent.

Let z_j = no. of the ball drawn in jth draw.

Then sample space S is

$$S = \{(z_1, z_2, \ldots\ldots z_n)\}$$

and $$O(S) = M^n$$

Let A_k = The event that there are exactly k defective balls.

Then A_k = Subset of S for which exactly k of the z_j's are numbers s.t $1 \le z_j \le K$ and the remaining $(n-k)$ z_j's are s.t $K+1 \le z_j \le M$.

Now in n-places, k–z_j's $(1 \le z_j \le K)$ can be arranged in ${}^n c_k$ ways; k such z_j's can be chosen in K^k ways and remaining $(n-k)$ z_j's can be chosen in $(M-K)^{n-k}$ ways.

$$\therefore \quad O(A_k) = {}^n c_k \cdot K^k \cdot (M-K)^{n-k}$$

$$\therefore \quad P(A_k) = \frac{{}^n c_k \cdot K^k \, (M-K)^{n-k}}{M^n}$$

(ii) *Without replacement*

Here $$O(S) = M(M-1)\ldots.(M-\overline{n-1}) = (M)_n$$

and $$O(A_k) = {}^n c_k \, (K)_k \, (M-K)_{n-k}$$

$$\therefore \quad P(A_k) = \frac{{}^n c_k (K)_k \, (M-K)_{n-k}}{(M)_n}$$

$$= \frac{{}^K c_k \cdot {}^{M-K} c_{n-k}}{{}^M c_n}$$

Ex. 7-118. *An urn contains M balls of which K are black and M–K are white. A sample of size n is drawn. Find the probability that the j^{th} ball drawn is black given that the sample contains k black balls.*

Sol. Two possibilities are there :

(i) sampling with replacement

(ii) sampling without replacement

Let A_k = Event that sample contains exactly k black balls.

B_j = Event that jth ball drawn is black

(i) *With replacement*

Here $$P(A_k) = {}^n c_k \cdot \frac{K^k (M-K)^{n-k}}{M^n} \qquad \text{(See Exercise 117)}$$

$$P(A_k / B_j) = {}^{n-1} c_{k-1} \cdot \frac{K^{k-1} \, (M-K)^{n-k}}{M^{n-1}}$$

$$P(B_j) = \frac{K}{M} \qquad (\because \text{ balls are replaced})$$

$$\therefore \quad P(B_j / A_k) = \frac{P(A_k / B_j) \cdot P(B_j)}{P(A_k)}$$

$$= \frac{\left\{ \dfrac{{}^{n-1}c_{k-1}\, K^{k-1}\, (M-K)^{n-k}}{M^{n-1}} \cdot \dfrac{K}{M} \right\}}{\dfrac{{}^{n}c_k\, K^k\, (M-K)^{n-k}}{M^n}}$$

$$= k/n.$$

(*ii*) *Without replacement*

Here
$$P(A_k) = {}^{K}C_k \cdot {}^{M-K}C_{n-k} \Big/ {}^{M}C_n$$

$$P(A_k / B_j) = {}^{K-1}C_{k-1} \quad {}^{M-K}C_{n-k} \Big/ {}^{M-1}C_{n-1}$$

$$P(B_j) = \sum_{i=0}^{j-1} P(B_j C_i) = \sum_{i=0}^{j-1} P(C_i)\, P(B_j / C_i)$$

where C_i = event of exactly i black balls in the first $(j-1)$ draws.

$$P(C_i) = {}^{K}C_i \cdot {}^{M-K}C_{j-1-i} \Big/ {}^{M}C_{j-1}$$

$$P(B_j / C_i) = \frac{K-i}{M-(j-1)}$$

This is because at jth draw, total no. of balls in the urn is $M-(j-1)$ out of which $K-i$ balls are black.

$$\therefore \quad P(B_j) = \sum_{i=0}^{j-1} \left\{ \frac{{}^{K}C_i\ {}^{M-K}C_{j-1-i}}{{}^{M}C_{j-1}} \right\} \cdot \frac{K-i}{M-j+1}$$

$$= \sum_{i=0}^{j-1} \left\{ \frac{\dfrac{K!}{i!\,(K-i)!}\ {}^{M-K}C_{j-1-i}}{\dfrac{M!}{(j-1)!\,(M-j+1)!}} \cdot \frac{K-i}{M-j+1} \right\}$$

$$= \frac{K}{M} \sum_{i=0}^{J-1} \left[\frac{(K-1)!}{i!\,(K-1-i)!} \cdot \frac{{}^{M-K}C_{j-1-i}}{{}^{M-1}C_{j-1}} \right]$$

$$= \frac{K}{M} \sum_{i=0}^{j-1} \left[\frac{{}^{K-1}C_i\ {}^{M-K}C_{j-1-i}}{{}^{M-1}C_{j-1}} \right]$$

P(L

∴ P(L

1. Prove or disprove :
   (*i*) If $P(A/B) \geq P($
   (*ii*) If $P(A) = P(\overline{B})$
   (*iii*) If $P(A) = 0$, th
   (*iv*) If $P(A) = P(B)$
   (*v*) If $P(B/\overline{A}) = P$
2. If A_1, A_2, B are events
   (*i*) $P(A_1 / B_2) = P($
   (*ii*) $P(A_1 / B) \leq P($
3. Show that
4. Examine the consistenc
5. *A*, *B* and *C* are three in

P(A

Prove that $P(A) = \dfrac{1}{2}$ a

$$= \frac{K}{M} \cdot \frac{\left({}^{M-1}C_{j-1}\right)}{{}^{M-1}C_{j-1}} = \frac{K}{M}.$$

$$P(B_j / A_k) = \frac{P(A_k / B_j)\, P(B_j)}{P(A_k)}$$

$$= \frac{\left\{ {}^{K-1}C_{k-1} \cdot {}^{M-K}C_{n-k} \Big/ {}^{M-1}C_{n-1} \right\} \cdot \left(K/M \right)}{{}^{K}C_{k} \cdot {}^{M-K}C_{n-k} \Big/ {}^{M}C_{n}}$$

$$= \frac{k}{n}$$

$$\therefore \qquad P(B_j / A_k) = \frac{\left\{ {}^{K-1}C_{k-1}\; {}^{M-K}C_{n-k} \Big/ {}^{M-1}C_{n-1} \right\} \cdot \frac{K}{M}}{{}^{K}C_{k} \cdot {}^{M-K}C_{n-k} \Big/ {}^{M}C_{n}} = \frac{k}{n}.$$

EXERCISES

1. Prove or disprove :

   (*i*) If $P(A / B) \geq P(A)$, then $P(B / A) \geq P(B)$

   (*ii*) If $P(A) = P(\overline{B})$ then $A = \overline{B}$

   (*iii*) If $P(A) = 0$, then $P(AB) = 0$

   (*iv*) If $P(A) = P(B) = p$, then $P(AB) \leq p^2$

   (*v*) If $P(B / \overline{A}) = P(B / A)$, then A and B are independent.

2. If A_1, A_2, B are events and $P(B) > 0$ show that :

   (*i*) $P(A_1 / B_2) = P(A_1 A_2 / B) + P(A_1 \overline{A}_2 / B)$

   (*ii*) $P(A_1 / B) \leq P(A_2 / B)$, provided $A_1 \subset A_2$.

3. Show that

$$P(A - B) = P(A) - P(A \cap B)$$

4. Examine the consistency of the following data

$$P(A) = \frac{1}{4},\ P(B) = \frac{1}{2},\ P(AB) = \frac{1}{8}.$$

5. A, B and C are three independent events such that

$$P(A\overline{B}\,\overline{C}) = \frac{1}{4},\ P(\overline{A}B\overline{C}) = \frac{1}{8},\ P(\overline{A}\,\overline{B}C) = \frac{1}{12}$$

   Prove that $P(A) = \dfrac{1}{2}$ and $P(\overline{A}\,\overline{B}\,\overline{C}) = \dfrac{1}{4}$.

6. Define a random variable. A random variable x takes values 0, 1, 2,.... with probability prcportional to $\frac{x+1}{5^x}$.

Find the prob. that $x \leq 5$.

7. From a pack of 52 cards, 13 cards are drawn without replacement. Find the probability that there are exactly 6 spade cards.

$$\left\{\textbf{Ans.:} \frac{^{13}C_6 \cdot {}^{39}C_7}{^{52}C_{13}}\right\}$$

8. If three squares are chosen at random on a chess board, show that the chance that they should be in a diagonal line is 7/744.

9. Three squares of a chess board being chosen at random, what is the chance that two are of one colour and one of another ?

$$\left[\textbf{Ans.} \frac{16}{21}\right]$$

10. A person writes 4 letters and 4 envelopes. If the letters are placed in the envelopes at random, what is the chance that not more than one letter is placed in the correct envelope ?

$$\left[\textbf{Ans.} \frac{17}{24}\right]$$

11. Four right-foot shoes are paired at random with the corresponding set of the left-foot shoes. Find the prob. that no correct pair is obtained.

$$\left[\textbf{Ans.} \frac{3}{8}\right]$$

12. From a pack of 52 cards, three are drawn at random. Find the chance that these are a king, a queen and a knave.

$$\left[\textbf{Ans.} \frac{16}{5525}\right]$$

13. If two balls are drawn from a bag containing 2 white, 4 red and 5 black balls. What is the chance that (*i*) both the balls are red, (*ii*) one is red and the other black ?

$$\left[\textbf{Ans.} \frac{6}{55}; \frac{4}{11}\right]$$

14. Find the chance of throwing a sum of 9 in a single throw of two dice.

$$\left[\textbf{Ans.} \frac{1}{9}\right]$$

15. Find the prob. of obtaining a total of 6 in a throw of 6 dice.

$$\left[\textbf{Ans.} \frac{1}{6^6}\right]$$

16. In a single throw of three
than 11, (*iii*) more than 1

17. A and B throw with 3 d
number ?

18. Find the chance of throw

19. Find the chance of throw

20. A person throws two dic
the number on the lowes
chance that the sum of th

21. There are 10 tickets 5 of
1, 2, 3, 4, 5. What is the p
replaced at every trial, (*ii*

22. Out of 20 consecutive nu
their sum is odd.

23. A bag contains 50 ticket
arranged in ascending or
probability that $x_3 = 30$?

24. Nine cards are drawn at r
the numbers '1', '0' or '–1
drawn. Find the chance th

ues 0, 1, 2,.... with probability

acement. Find the probability

$$\left\{\textbf{Ans.:} \frac{{}^{13}C_6 \cdot {}^{39}C_7}{{}^{52}C_{13}}\right\}$$

how that the chance that they

vhat is the chance that two are

$$\left[\textbf{Ans.}\ \frac{16}{21}\right]$$

re placed in the envelopes at
tter is placed in the correct

$$\left[\textbf{Ans.}\ \frac{17}{24}\right]$$

esponding set of the left-foot

$$\left[\textbf{Ans.}\ \frac{3}{8}\right]$$

d the chance that these are a

$$\left[\textbf{Ans.}\ \frac{16}{5525}\right]$$

ed and 5 black balls. What is
id the other black ?

$$\left[\textbf{Ans.}\ \frac{6}{55}; \frac{4}{11}\right]$$

of two dice.

$$\left[\textbf{Ans.}\ \frac{1}{9}\right]$$

e.

$$\left[\textbf{Ans.}\ \frac{1}{6^6}\right]$$

16. In a single throw of three dice, what is the chance of throwing (*i*) 'four-five-six, (*ii*) less than 11, (*iii*) more than 10 ?'

$$\left[\textbf{Ans.}\ \frac{1}{36}; \frac{1}{2}; \frac{1}{2}\right]$$

17. A and B throw with 3 dice; if A throws 8, what is B's chance of throwing a higher number ?

$$\left[\textbf{Ans.}\ \frac{20}{27}\right]$$

18. Find the chance of throwing 10 exactly in one throw with 3 dice.

$$\left[\textbf{Ans.}\ \frac{1}{8}\right]$$

19. Find the chance of throwing (*i*) 18, (*ii*) 10 exactly in one throw of 4 dice.

$$\left[\textbf{Ans.}\ \frac{5}{81}; \frac{5}{81}\right]$$

20. A person throws two dice, one the common cube and the other a regular tetrahedron, the number on the lowest face being taken in the case of the tetrahedron; what is the chance that the sum of the numbers thrown is not less than 5 ?

$$\left[\textbf{Ans.}\ \frac{3}{4}\right]$$

21. There are 10 tickets 5 of which are blanks and the others are marked with the numbers 1, 2, 3, 4, 5. What is the prob. of drawing 10 in the three trials, (*i*) when the tickets are replaced at every trial, (*ii*) if the tickets are not replaced ?

$$\left[\textbf{Ans.}\ \frac{33}{1000}, \frac{1}{60}\right]$$

22. Out of 20 consecutive numbers two are chosen at random, find the probability that their sum is odd.

$$\left[\textbf{Ans.}\ \frac{10}{19}\right]$$

23. A bag contains 50 tickets numbered 1, 2,...50 of which 5 are drawn at random and arranged in ascending order of their numbers. $(x_1 < x_2 < x_3 < x_4 < x_5)$. What is the probability that $x_3 = 30$?

$$\left[\textbf{Ans.}\ \frac{{}^{29}c_2 \times {}^{20}c_2}{{}^{50}c_5}\right]$$

24. Nine cards are drawn at random from a set of cards. Each card is marked with one of the numbers '1', '0' or '–1' and it is equally likely that any of the three numbers will be drawn. Find the chance that the sum of the numbers drawn is zero.

$$\left[\textbf{Ans.}\ \frac{3139}{3^9}\right]$$

25. A party of 21 persons take their seats at a round table. What are the odds in favour of two specified persons sitting together ?

[Ans. 1 : 9]

26. A number is chosen from each of two sets :
1, 2, 3, 4, 5, 6, 7, 8, 9; 1, 2, 3, 4, 5, 6, 7, 8, 9.
If P_1 denotes the probability that the sum of the numbers be 10 and P_2 the probability that their sum be 8, find $P_1 + P_2$.

$$\left[\textbf{Ans. } \frac{16}{81}\right]$$

27. A card is drawn from an ordinary pack and a gambler bets that it is a spade or an ace. What are the odds against his winning his bet ?

[Ans. 9 : 4]

28. Find the probability that in a random arrangement of the letters of the word 'UNIVERSITY' the two 'I's don't come together.

29. A party of 'n' men of whom 'A', 'B' are two form single rank. What is the chance that (*i*) A, B are next one another, (*ii*) exactly 'm' men are between them, (*iii*) not more than 'm' men are between them ?

$$\left[\textbf{Ans. } \frac{2}{n};\ \frac{2(n-m-1)}{n(n-1)};\ \frac{(m+1)(2n-m-2)}{n(n-1)}\right]$$

30. If the letters of 'ATTEMPT' are written down at random, find the chance that (*i*) all the 'T's are together, (*ii*) no two 'T's are together.

$$\left[\textbf{Ans. } \frac{1}{7};\ \frac{2}{7}\right]$$

31. Find the number of ways in which 'p plus signs' and 'q minus signs' may be placed in a row so that no two minus signs are together.

32. A letter is chosen at random out of 'ASSININE' and one is chosen at random out of 'ASSASSIN'. Show that the chance that the same letter is chosen on both occasions is $\frac{1}{4}$.

33. Six cards are drawn at random from a pack of 52 cards. What is the probability that 3 will be red and 3 black ?

$$\left[\textbf{Ans. } \frac{13000}{39151}\right]$$

34. A bag contains 6 white and 9 black balls. The drawings of 4 balls are made such that (*a*) the balls are replaced before the second draw, (*b*) the balls are not replaced before the second draw. Find the probability that the first drawing will give 4 white and the second 4 black balls in each case.

$$\left[\textbf{Ans. } \frac{6}{5915};\ \frac{3}{715}\right]$$

35. Obtain the probability tha
days of the week, assumin

36. From a group of 25 persor
(Assume a 365-day year a

37. Cards are drawn one-by-o
cards will precede the firs

(Hint: See 7-93)

38. The odds against A solving
the same problem are 7 to
both try ?

39. A hand of 13 cards is deal
that the hand contains no
contains 4 cards of one su

40. Two drawings each of 3 ba
the balls being replaced be
will give 3 white and the s

41. A and B draw from a bag
chances of first drawing a

42. A is one of 6 horses entere
C. It is 2 : 1 that B rides A
rides A his chance is triple

43. A person draws a card fro
doing so until he draws a 'c
three trials, (*ii*) exactly thr

44. Three urns contain respecti
2 white and 3 black balls. C
among the balls drawn the

45. A bag contains 17 counter
replaced; a second drawing

Vhat are the odds in favour of

[**Ans.** 1 : 9]

s be 10 and P_2 the probability

$\left[\textbf{Ans. } \frac{16}{81}\right]$

ets that it is a spade or an ace.

[**Ans.** 9 : 4]

t of the letters of the word

e rank. What is the chance that
ween them, (*iii*) not more than

$\frac{-m-1)}{n-1)}; \frac{(m+1)(2n-m-2)}{n(n-1)}\Big]$

ı, find the chance that (*i*) all the

$\left[\textbf{Ans. } \frac{1}{7}; \frac{2}{7}\right]$

minus signs' may be placed in

one is chosen at random out
me letter is chosen on both

. What is the probability that 3

$\left[\textbf{Ans. } \frac{13000}{39151}\right]$

s of 4 balls are made such that
e balls are not replaced before
ving will give 4 white and the

$\left[\textbf{Ans. } \frac{6}{5915}; \frac{3}{715}\right]$

35. Obtain the probability that the birth-days of seven people will fall on seven different days of the week, assuming equal probability for the seven days.

$\left[\textbf{Ans. } \frac{6!}{7^6}\right]$

36. From a group of 25 persons, what is the prob. that all 25 will have different **birthdays**. (Assume a 365-day year and that all days are equally likely).

37. Cards are drawn one-by-one from a full deck. What is the probability that **exactly 10** cards will precede the first ace.

(**Hint:** See 7-93) $\left[\textbf{Ans. } \frac{164}{4165}\right]$

38. The odds against A solving a certain problem are 4 to 3 and odds in favour of B solving the same problem are 7 to 5. What is the probability that the problem is solved if they both try ?

$\left[\textbf{Ans. } \frac{16}{21}\right]$

39. A hand of 13 cards is dealt out randomly from a full deck of 52 cards. Find the prob. that the hand contains no spade card. Also show that the probability that the hand contains 4 cards of one suit and 3 cards each of the other three suits is

$$4.{}^{13}c_4.({}^{13}c_3)^3 \Big/ {}^{52}c_{13}.$$

40. Two drawings each of 3 balls are made from a bag containing 5 white and 8 black balls, the balls being replaced before the second trial. Find the chance that the first drawing will give 3 white and the second 3 black balls.

$\left[\textbf{Ans. } \frac{140}{20449}\right]$

41. A and B draw from a bag containing 3 white and 4 black balls. Find their respective chances of first drawing a white ball (the balls when drawn not being replaced).

$\left[\textbf{Ans. } \frac{22}{35}, \frac{13}{35}\right]$

42. A is one of 6 horses entered for a race and is to be ridden by one of two jockeys B and C. It is 2 : 1 that B rides A in which case all the horses are equally likely to win; if C rides A his chance is tripled. What are the odds against his winning ? [**Ans. 13 : 5**]

43. A person draws a card from a pack, replaces it, and shuffles the pack. He continues doing so until he draws a 'club'. What is the chance that he will have to make (*i*) at least three trials, (*ii*) exactly three trials ?

$\left[\textbf{Ans. } \frac{9}{16}, \frac{9}{64}\right]$

44. Three urns contain respectively 1 white and 2 black balls; 3 white and 1 black balls and 2 white and 3 black balls. One ball is taken at random from each urn. Find the prob. that among the balls drawn there are 2 white and 1 black ball.

45. A bag contains 17 counters marked with the numbers 1 to 17. A counter is drawn and replaced; a second drawing is then made, find the chance that the first number drawn is

even and the second odd. $\left[\textbf{Ans. } \dfrac{72}{289}\right]$

46. Three urns respectively contain 1 white and 3 black, 2 white and 4 black and 3 white and 1 black balls. A ball is drawn from an urn selected at random, find the chance of its being white.

$\left[\textbf{Ans. } \dfrac{4}{9}\right]$

47. Criticise the statement : 'the chance of throwing ace in the first trial is $\dfrac{1}{6}$ and the chance of ace in the second trial is $\dfrac{1}{6}$, therefore the chance of ace in two trials is $\dfrac{1}{3}$.'

48. Counters marked 1, 2, 3 are placed in a bag and one is withdrawn and replaced. The operation being repeated three times, what is the chance of obtaining a total of 6 ?

$\left[\textbf{Ans. } \dfrac{7}{27}\right]$

49. A, B, C in order cut a pack of cards, replacing them after each cut, on the condition that the first who cuts a spade shall win a prize. Find their respective chances.

$\left[\textbf{Ans. } \dfrac{16}{37}; \dfrac{12}{37}; \dfrac{9}{37}\right]$

50. Six persons throw for a stake, which is to be won by the one who first throws head with a coin. If they throw in succession, find the chance of the fourth person.

$\left[\textbf{Ans. } \dfrac{4}{63}\right]$

51. A and B play for a prize; A is to throw a die first and is to win if he throws 6. If he fails, B is to throw and to win if he throws 6 or 5. If he fails, A is to throw again and to win with 6 or 5 or 4, and so on. Find the chance of each player.

$\left[\textbf{Ans. } A\,\dfrac{169}{324}; B, \dfrac{155}{324}\right]$

52. A certain stake is to be won by the first person who throws an ace with an octahedral die. If there are 4 persons, what is the chance of the last ?

$\left[\textbf{Ans. } \dfrac{343}{1695}\right]$

53. A, B, C, D cut a pack of cards successively in the order mentioned. What are their respective chances of first cutting a spade ?

$\left[\textbf{Ans. } \dfrac{64}{175}; \dfrac{48}{175}; \dfrac{36}{175}; \dfrac{27}{175}\right]$

54. Five persons A, B, C, D, E throw a die in the order named until one of them throws an ace; find their respective chances of winning, supposing the throws to continue till an ace appears.

$\left[\textbf{Ans. } 1 : \dfrac{5}{6} : \left(\dfrac{5}{6}\right)^2 : \left(\dfrac{5}{6}\right)^3 : \left(\dfrac{5}{6}\right)^4\right]$

55. How many throws with a p

getting 'double six' at least

56. How many throws with a sin

an ace at least once greater

57. If the prob. of success be 0·

least one success is greater

58. The odds against a certain
independent of the former, a
happen.

59. The odds that a book will be
4 : 3 and 3 : 4 respectively.
will be favourable ?

60. A throws two coins and B
greater number of heads tha

61. There are three works, one
volume. They are placed on

same works are all together

62. There are two bags, one of
and 12 white balls. One ball
change of drawing a red bal

63. Four students are selected at
that the selected students in
least two boys.

64. A bag contains 4 white, 5 re
prob. that
(*a*) No ball drawn is black.
(*b*) Exactly two are black.
(*c*) All are of the same colo

$\left[\textbf{Ans. } \frac{72}{289}\right]$

white and 4 black and 3 white
ıt random, find the chance of its

$\left[\textbf{Ans. } \frac{4}{9}\right]$

in the first trial is $\frac{1}{6}$ and the
ıance of ace in two trials is $\frac{1}{3}$.'
s withdrawn and replaced. The
:e of obtaining a total of 6 ?

$\left[\textbf{Ans. } \frac{7}{27}\right]$

r each cut, on the condition that
respective chances.

$\left[\textbf{Ans. } \frac{16}{37}; \frac{12}{37}; \frac{9}{37}\right]$

one who first throws head with
:he fourth person.

$\left[\textbf{Ans. } \frac{4}{63}\right]$

o win if he throws 6. If he fails,
A is to throw again and to win
ıyer.

$\left[\textbf{Ans. } A\frac{169}{324}; B, \frac{155}{324}\right]$

rows an ace with an octahedral
ıt ?

$\left[\textbf{Ans. } \frac{343}{1695}\right]$

der mentioned. What are their

$\left[\textbf{Ans. } \frac{64}{175}; \frac{48}{175}; \frac{36}{175}; \frac{27}{175}\right]$

ıed until one of them throws an
ıg the throws to continue till an

$1 : \frac{5}{6} : \left(\frac{5}{6}\right)^2 : \left(\frac{5}{6}\right)^3 : \left(\frac{5}{6}\right)^4\Big]$

55. How many throws with a pair of dice are necessary in order to have the chance of getting 'double six' at least once greater than $\frac{1}{2}$?

[Ans. 25]

56. How many throws with a single die are necessary in order to have the chance of getting an ace at least once greater than $\frac{1}{2}$? **[Ans. 4]**

57. If the prob. of success be 0·01, how many trials are necessary in order that prob. of at least one success is greater than $\frac{1}{2}$. **[Ans. 69]**

58. The odds against a certain event are 5 : 2 and the odds in favour of another event, independent of the former, are 6 : 5. Find the chance that one at least of the events will happen.

$\left[\textbf{Ans. } \frac{52}{77}\right]$

59. The odds that a book will be favourably reviewed by three independent critics are 5 : 2, 4 : 3 and 3 : 4 respectively. What is the probability that of the three reviews a majority will be favourable ?

$\left[\textbf{Ans. } \frac{209}{343}\right]$

60. *A* throws two coins and *B* throws three coins. Find the chance that *B* will throw a greater number of heads than *A*.

$\left[\textbf{Ans. } \frac{1}{2}\right]$

61. There are three works, one consisting of '3' volumes, one of '4' and the other of '1' volume. They are placed on a self at random; prove that the chance that volumes of the same works are all together is $\frac{3}{140}$.

62. There are two bags, one of which contains 5 red and 7 white balls and the other 3 red and 12 white balls. One ball is to be drawn from one or other of the two bags. Find the change of drawing a red ball.

$\left[\textbf{Ans. } \frac{37}{120}\right]$

63. Four students are selected at random from 7 boys and 4 girls. Calculate the probabilities that the selected students include (*i*) two specified boys, (*ii*) exactly two boys, (*iii*) at least two boys.

64. A bag contains 4 white, 5 red and 6 black balls. Three are drawn at random. Find the prob. that

(*a*) No ball drawn is black.
(*b*) Exactly two are black.
(*c*) All are of the same colour.

65. A has '3' shares in a lottery in which there are '3' prizes and '6' blanks B has '1' share in a lottery in which there is '1' prize and '2' blanks. Show that A's chance of success is to B's as 16 : 7.

66. If p is the prob. that a man aged x years will die in a year. Find the prob. that out of 'm' men $A_1, A_2 \ldots A_m$, each aged x, A will die in a year and be the first to die.

$$\left[\textbf{Ans. } \frac{1}{m}\{1-(1-p)^m\}\right]$$

67. It is 8 : 5 against a person who is now 40 years old living till he is 70 and 4 : 3 against a person now 50 living till he is 80. Find the prob. that at least one of these persons will be alive 30 years hence. $\left[\textbf{Ans. } \frac{59}{91}\right]$

68. The prob. that a 50 years old man will be alive at 60 is 0·83 and the prob. that a 45 years old woman will be alive at 55 is 0·87. What is the prob. that a man who is 50 and his wife who is 45 will both be alive 10 year hence ?

[**Ans.** 0·7221]

69. Suppose that it is 9 : 7 against a person A who is now 35 years of age living till he is 65 and 3 : 2 against a person B now 45 living till he is 75; find the chance that one at least of these persons will be alive 30 years hence.

$$\left[\textbf{Ans. } \frac{53}{80}\right]$$

70. A number consists of 7 digits whose sum is 59; prove that the chance of its being divisible by 11 is $\frac{4}{21}$.

71. If two coins are tossed 5 times, what is the chance that there will be 5 heads and 5 tails ? $\left[\textbf{Ans. } \frac{63}{256}\right]$

72. Find the chance of obtaining at least one six in a throw of four dice.

73. Show that the chance of throwing at least one ace in a single throw with two dice is $\frac{11}{36}$.

74. What is the probability of getting 9 cards of the same suit in one hand at a game of bridge ? $\left[\textbf{Ans. } \frac{{}^{13}C_9\ {}^{39}C_4 . {}^{4}C_1}{{}^{52}C_{13}}\right]$

75. In three throws with a pair of dice, find the chance of throwing doublets at least once.

$$\left[\textbf{Ans. } \frac{91}{216}\right]$$

76. One bag contains 3 white balls and 2 black balls, another contains 5 white and 3 black balls. If a bag is chosen at random and a ball is drawn from it, what is the chance that it is white ? $\left[\textbf{Ans. } \frac{49}{80}\right]$

77. Four cards are drawn witho

78. Find prob. in Ex. 40 if the t

79. If the war breaks out on the
a stretch there will be no w

80. Three newspapers A, B, C a
that of the adult population
B, 5% read both A and C, 4(
read at least one of the pape
both A and B ?

81. There are three boxes cont
3 red, 1 black balls; 3 whit
it two balls are drawn at ra
prob. that they come from

82. The probability that a persc
hit the same target is $\frac{2}{5}$. I
person fires 5 shots. They f
the target ?

s and '6' blanks B has '1' share
how that A's chance of success

r. Find the prob. that out of 'm'
l be the first to die.

$$\left[\textbf{Ans. } \frac{1}{m}\{1-(1-p)^m\}\right]$$

ıg till he is 70 and 4 : 3 against
t least one of these persons will

$$\left[\textbf{Ans. } \frac{59}{91}\right]$$

83 and the prob. that a 45 years
). that a man who is 50 and his

[**Ans.** 0·7221]

; years of age living till he is 65
ind the chance that one at least

$$\left[\textbf{Ans. } \frac{53}{80}\right]$$

'e that the chance of its being

at there will be 5 heads and 5

$$\left[\textbf{Ans. } \frac{63}{256}\right]$$

of four dice.
a single throw with two dice

suit in one hand at a game of

$$\left[\textbf{Ans. } \frac{{}^{13}c_9\ {}^{39}c_4 . {}^{4}c_1}{{}^{52}c_{13}}\right]$$

ırowing doublets at least once.

$$\left[\textbf{Ans. } \frac{91}{216}\right]$$

er contains 5 white and 3 black
om it, what is the chance that it

$$\left[\textbf{Ans. } \frac{49}{80}\right]$$

77. Four cards are drawn without replacement. What is the prob. that these are all aces ?

$$\left[\textbf{Ans. } \frac{1}{270725}\right]$$

78. Find prob. in Ex. 40 if the balls are not replaced before the second draw.

$$\left[\textbf{Ans. } \frac{7}{429}\right]$$

79. If the war breaks out on the average once in 25 years, find the prob. that in 50 years at a stretch there will be no war.

$$\left[\textbf{Ans. } e^{-2}\right]$$

80. Three newspapers A, B, C are published in a certain city. It is estimated from a survey that of the adult population : 20% read A, 16% read B, 14% read C, 8% read both A and B, 5% read both A and C, 40% read both B and C, 2% read all three (*i*) What percentage read at least one of the papers ? (*ii*) of those that read at least one, what percentage read both A and B ?

[**Ans.** 35%, 28%]

81. There are three boxes containing respectively 1 white, 2 red, 3 black balls; 2 white, 3 red, 1 black balls; 3 white, 1 red, 2 black balls. A box is chosen at random and from it two balls are drawn at random. The two balls are one red and one white. What is the prob. that they come from the (*i*) 1st box, (*ii*) 2nd box, (*iii*) 3rd box ?

$$\left[\textbf{Ans. } \frac{2}{11}, \frac{6}{11}, \frac{3}{11}\right]$$

82. The probability that a person can hit a target is $\frac{3}{5}$ and the prob. that another person can hit the same target is $\frac{2}{5}$. But the first person can fire 4 shots in the time the second person fires 5 shots. They fire together. What is the prob. that the second person shoots the target ?

$$\left[\textbf{Ans. } \frac{5}{11}\right]$$

8

Mathematical Expectation

8.1. Stochastic Variate. *The variate which can take certain values depending on chance is called* **chance variate** *or* **stochastic variate** *or* **random variate** *e.g., In rolling a die the variate corresponding to the number obtained is a stochastic variate.*

Variables are generlly denoted by capital letters (*i.e.*, X, Y etc.) and corresponding small letters represents their values. In certain cases some small letters are used for both purposes.

Probability Distribution. *The dist obtained by taking the possible values of a chance variate together with their respective probabilities is called prob. dist.*

Expected value of a chance variate.

Let x be the chance variate with prob dist.

$$\begin{array}{l} x \to \\ p \to \end{array}\begin{pmatrix} x_1 & x_2 \ldots\ldots\ldots x_n \\ p_1 & p_2 \ldots\ldots\ldots p_n \end{pmatrix}$$

Then expected value of x is defined to be $x_1p_1 + x_2p_2 + \ldots + x_np_n$ and is denoted by $E(x)$. Thus, $E(x) = x_1p_1 + x_2p_2 + \ldots + x_np_n$.

Ex. 8-1. *Let x denote the profit that a man makes in a business. He may earn 2,800 with probability 0·5; he may lose Rs. 5,500 with probability 0·3 and he may neither earn nor lose with probability 0·2. Calculate the mathematical expectation of x.*

Sol. The given prob. dist is

$$\begin{array}{l} x \to \\ p \to \end{array}\begin{pmatrix} -5500 & 0 & 2800 \\ 0\cdot3 & 0\cdot2 & 0\cdot5 \end{pmatrix}$$

$$\therefore \; E(x) = (-5500)(0\cdot3) + (2800)(0\cdot5)$$

$$= -1650 + 1400 = -250.$$

Ex. 8-2. *Find the expected value of the number of points that will be obtained in a single throw with an ordinary die.*

Sol. Lex x be the number of points obtained in a single throw with an ordinary dice. Then x can take values 1, 2, 3, 4, 5, 6.

Also prob. of getting any number with a single die

$$= \frac{1}{6}$$

Therefore, expected value of x

$$= \frac{1}{6}\{1+2+3+4+5+6\}$$

$$= \frac{21}{6}$$

$$= \frac{7}{2}.$$

Ex. 8-3. *From a bag contai*
draw 2 coins indiscriminately. F
Sol. Prob. of drawing 2 '20

Prob. of drawing 1 '20 P' c

Prob. of drawing 2 '25 P' c

The person gets 40 P, 45 P

∴ Expectation of the perso

Ex. 8-4. *A person draws 2 b*
to receive 10 P for every white
expectation.
Sol. Three different possibi
(*i*) The person draws 2 white

(*ii*) The person draws 1 whi
this happening

(*iii*) The person draws 2 red

∴ Expectation

Ex. 8-3. *From a bag containing 2 '20P' coins and 3 '25 P' coins, a person is allowed to draw 2 coins indiscriminately. Find the value of his expectation.*

Sol. Prob. of drawing 2 '20 P' coins

$$= \frac{{}^2c_2}{{}^5c_2} = \frac{1}{10}$$

Prob. of drawing 1 '20 P' coin and 1 '25 P' coin

$$= \frac{{}^2c_1 \cdot {}^3c_1}{{}^5c_2} = \frac{3}{5}$$

Prob. of drawing 2 '25 P' coins

$$= \frac{{}^3c_2}{{}^5c_2} = \frac{3}{10}.$$

The person gets 40 P, 45 P and 50 P in three cases respectively.

$$\therefore \text{ Expectation of the person } = 40\frac{1}{10} + 45\frac{3}{5} + 50\frac{3}{10}$$

$$= 4 + 27 + 15 = 46\text{P}$$

Ex. 8-4. *A person draws 2 balls from a bag containing 3 white and 4 red balls. If he is to receive 10 P for every white ball which he draws and 20 P for each red ball. Find his expectation.*

Sol. Three different possibilities are :

(*i*) The person draws 2 white balls. In this case he gets 20 P and the prob. of this happening

$$= \frac{{}^3c_2}{{}^7c_2} = \frac{3}{21}$$

(*ii*) The person draws 1 white and 1 red ball. In this case he gets 30 P and the prob. of this happening

$$= \frac{{}^3c_1 \times {}^4c_1}{{}^7c_2} = \frac{12}{21}.$$

(*iii*) The person draws 2 red balls. In this case he gets 40 P and the prob. of this happening

$$= \frac{{}^4c_2}{{}^7c_2} = \frac{6}{21}$$

$$\therefore \text{ Expectation } = \frac{1}{21}\{20.3 + 30.12 + 40.6\}$$

$$= \frac{1}{21}\{60 + 360 + 240\}$$

$$= \frac{1}{21}(660)$$

$$= \frac{220}{7} \simeq 31\text{ P.}$$

Ex. 8-5. *Three urns contain respectively 3 green and 2 white balls, 5 green and 6 white balls and 2 green and 4 white balls. One ball is drawn from each urn. Find the expected number of white balls drawn out.*

Sol. Let x be the number of white balls drawn. Then possible values of x are 0, 1, 2 and 3.

Let p_0, p_1, p_2 and p_3, be the probabilities of x taking these values respectively.

Now p_0 = prob. of drawing all the three green balls.

$$= \frac{3}{5}\cdot\frac{5}{11}\cdot\frac{2}{6} = \frac{1}{11}.$$

Different possibilities of drawing 1 white and 2 green balls are :

1st urn	2nd urn	3rd urn
W	G	G
G	W	G
G	G	W

Where 'G' denotes the green ball and 'W' the white ball.

$$\therefore \qquad p_1 = \frac{2}{5}\cdot\frac{5}{11}\cdot\frac{2}{6} + \frac{3}{5}\cdot\frac{6}{11}\cdot\frac{2}{6} + \frac{3}{5}\cdot\frac{5}{11}\cdot\frac{4}{6} = \frac{58}{165}$$

Similarly,

$$p_2 = \frac{3}{5}\cdot\frac{6}{11}\cdot\frac{4}{6} + \frac{2}{5}\cdot\frac{5}{11}\cdot\frac{4}{6} + \frac{2}{5}\cdot\frac{6}{11}\cdot\frac{2}{6}$$

$$= \frac{68}{165}$$

$$p_3 = \frac{2}{5}\cdot\frac{6}{11}\cdot\frac{4}{6} = \frac{8}{55}$$

$$\therefore \qquad E(x) = 0.\frac{1}{11} + 1.\frac{58}{165} + 2.\frac{68}{165} + 3.\frac{8}{55} = \frac{266}{165}.$$

Ex. 8-6. *What is the expectation of the number of failures preceding the first success in an indefinite series of independent trials with constant probability p of success ?*

Sol. Let x be the number of failures preceding the first success. Then x can take values

$$0, 1, 2 \ldots\ldots\ldots\ldots n \ldots\ldots\ldots\ldots$$

with respective probabilities

$$p, qp, q^2p, \ldots\ldots q^n P \ldots\ldots\ldots$$

$$\therefore \qquad E(x) = 0.p + 1.qp + 2.q^2p + \ldots + n.q^n p + \ldots$$

$$= pq(1 + 2q + \ldots\ldots + nq^{n-1} + \ldots\ldots)$$

$$= pq(1-q)^2$$

$$= \frac{pq}{p^2} = q/p.$$

Ex. 8-7. *A makes a bet with B of Rs. 5 to Rs. 2 that in a single throw with two dice he will throw 7 before B throws 4. Each has a pair of dice and they throw simultaneously until one of them wins, equal throws being disregarded. Find B's expectation.*

Sol. Prob. of getting 7 in a

and prob. of gett

Since total prob. is unity, p

Now A wins if he throws 7 l
not throw 7 or 4.

$\therefore$ Prob. of A winning in fi

and prob. of B winning in 1

$\therefore$ Prob. of none-winning i

Now A wins in second tri
neither 7 nor 4 in second trial.

$\therefore$ Prob. of A winning in se

Similarly prob. of B winni

prob. of A winning in thirc

prob. of B winning in thirc

and so on.

$\therefore$ A's chance of winning

Sol. Prob. of getting 7 in a single throw with two dice

$$= \frac{1}{6}$$

$$\text{and prob. of getting } 4 = \frac{1}{12}.$$

Since total prob. is unity, prob. of throwing neither 7 nor 4

$$= 1 - \frac{1}{6} - \frac{1}{12} = \frac{3}{4}.$$

Now A wins if he throws 7 but B does not throw 7 or 4 and B wins if throws 4 but A does not throw 7 or 4.

$\therefore$ Prob. of A winning in first trial

$$= \frac{1}{6} \cdot \frac{3}{4} = \frac{1}{8}$$

and prob. of B winning in first trial

$$= \frac{1}{12} \cdot \frac{3}{4} = \frac{1}{16}$$

$\therefore$ Prob. of none-winning in the first trial

$$= 1 - \frac{1}{8} - \frac{1}{16} = \frac{13}{16}$$

Now A wins in second trial, if in first trial none wins and he throws 7 but B throws neither 7 nor 4 in second trial.

$\therefore$ Prob. of A winning in second trial

$$= \frac{13}{16} \cdot \frac{1}{6} \cdot \frac{3}{4} = \frac{13}{16} \cdot \frac{1}{8}$$

Similarly prob. of B winning in second trial

$$= \frac{13}{16} \cdot \frac{1}{12} \cdot \frac{3}{4} = \frac{13}{16} \cdot \frac{1}{16}$$

prob. of A winning in third throw

$$= \left(\frac{13}{16}\right)^2 \cdot \frac{1}{8}$$

prob. of B winning in third throw

$$= \left(\frac{13}{16}\right)^2 \cdot \frac{1}{16}$$

and so on.

$\therefore$ A's chance of winning

$$= \frac{1}{8}\left\{1 + \frac{13}{16} + \left(\frac{13}{16}\right)^2 + \ldots\ldots\right\}$$

$$= \frac{1}{8} \cdot \frac{1}{1 - \frac{13}{16}} = \frac{2}{3}$$

and B's chance of winning

$$= \frac{1}{16}\left\{1+\frac{13}{16}+\left(\frac{13}{16}\right)^2+\ldots\ldots\right\}$$

$$= \frac{1}{16}\cdot\frac{1}{1-\frac{13}{16}}=\frac{1}{3}$$

Now A gets Rs. $2.\frac{2}{3}=\frac{4}{3}$, if he wins and pays Rs. $5.\frac{1}{3}=\frac{5}{3}$, if he loses.

$\therefore$ B's expectation $= \frac{5}{3}-\frac{4}{3}=$ Rs. $\frac{1}{3}$.

Ex. 8-8. *A coin is tossed until a head appears. What is the expectation of the number of tosses.*

Sol. Prob. of getting a head in a toss $=\frac{1}{2}=$ Prob. of getting a tail in a toss.

Let x be the number of tosses until a head appears. Then x can take values 1,2, 3,.........................

When x takes value 1, head appears in very first trial and the prob. for this is $\frac{1}{2}$. When x takes value 2, first trial results in tail and second in head. So by compound prob. theorem, prob. that x takes value $2=\left(\frac{1}{2}\right)^2$. Similarly prob. that x takes value $3=\left(\frac{1}{2}\right)^3$ and so on.

Therefore, expected value of x

$$= \frac{1}{2}.1+\left(\frac{1}{2}\right)^2.2+\left(\frac{1}{2}\right)^3.3+\ldots.$$

$$= \frac{1}{2}\left\{1+2.\frac{1}{2}+3.\left(\frac{1}{2}\right)^2+\ldots\ldots.\right\}$$

$$= \frac{1}{2}.\left(1-\frac{1}{2}\right)^{-2}$$

$$= \frac{\frac{1}{2}}{\frac{1}{4}}=2.$$

8.2. Indicator Function (for discrete variable)

For a random variable x indicator fuction is defined by

$$I_{(x_1,\,x_2,\ldots\ldots x_n)}(x) = \begin{cases}1 & \text{if } x=x_i\\ 0 & \text{if } x\neq x_i\end{cases} \quad (i=1,2\ldots\ldots n)$$

Discrete density functio

Let x be a random varial

$f_x(\cdot)$ of x is defined by

$f_x(\cdot)$ is called *discrete d*

$f_x(\cdot)$ is a function with c

Remark: (*i*) By using in

(*ii*) Cumulative distributi

and converse formula is

e.g., Let x denote the numbe

1, 2,.......6 and probability ass

$\therefore$

$F_x($

Then F

using etc. remark

and by using equation (3)

....}

$\frac{5}{3}$, if he loses.

expectation of the number of

ng a tail in a toss.

: can take values

ne prob. for this is $\frac{1}{2}$. When

y compound prob. theorem,

value $3 = \left(\frac{1}{2}\right)^3$ and so on.

....

....}

......n)

Discrete density function of a discrete random variable.

Let x be a random variable with distinct values $x_1, x_2, \ldots\ldots x_n$. The density function $f_x(\cdot)$ of x is defined by

$$f_x(x) = \begin{cases} P(x = x_i) & \text{if } x = x_i \\ 0 & \text{if } x \neq x_i \quad (i = 1, 2 \ldots\ldots n) \end{cases}$$

$f_x(\cdot)$ is called *discrete density function* of x.

$f_x(\cdot)$ is a function with domain real line and range the interval $[0, 1]$.

Remark: (*i*) By using indicator function,

$$f_x(x) = \sum_{i=1}^{n} P(x = x_i)\, I_{(x_i)}(x) \qquad \ldots(1)$$

(*ii*) Cumulative distribution function is given by

$$F_x(x) = \sum_{i: x_i \leq x} f_x(x_i) \qquad \ldots(2)$$

and converse formula is

$$f_x(x_i) = F_x(x_i) - \sum_{0<h\to 0} F_x(x_i - h) \qquad \ldots(3)$$

e.g., Let x denote the number obtained in rolling a single die. Possible values of x are $1, 2, \ldots\ldots 6$ and probability associated with each value is $\frac{1}{6}$.

$$\therefore \qquad f_x(x) = \frac{1}{6} I_{(1, 2, \ldots\ldots, 6)}(x)$$

or

$$\sum_{i=1}^{6} \frac{1}{6} I_{(i)}(x)$$

$$F_x(x) = \sum_{i=1}^{5} \frac{i}{6} I_{(i \leq x \leq i+1)}(x) + I_{(6 \leq x < \infty)}(x)$$

Then
$$F_x(3\cdot 5) = \sum_{i=1}^{5} \frac{i}{6} I_{(i \leq 3\cdot 5 \leq i+1)}(3\cdot 5)$$

$$= \frac{3}{6}.$$

using etc. remark

$$f_x(3) = P(x = 3) = \frac{1}{6} \qquad \text{[using equation (1)]}$$

and by using equation (3)

$$f_x(3) = F_x(3) - \sum_{0<h\to 0} F_x(3 - h)$$

$$= F_x(3) - F_x(2)$$

$$= \frac{3}{6} - \frac{2}{6} = \frac{1}{6}.$$

Ex. 8-9. *A coin is tossed until a head appears. Let x denote the number of tosses. Find*
(i) density function of x
(ii) mean and variable of x
(iii) moment generating function of x.

Sol. For (*i*) and mean see Ex. 8-8.

To find variance:

$$S = E(x^2) = 1^2.\frac{1}{2} + 2^2.\frac{1}{2^2} + 3^2.\frac{1}{2^3} + 4^2.\frac{1}{2^4} + \ldots\ldots\ldots$$

$$= 1.\frac{1}{2} + 4.\frac{1}{2^2} + 9\frac{1}{2^3} + 16.\frac{1}{2^4} + 25.\frac{1}{2^5} + \ldots\ldots\ldots$$

$$\frac{S}{2} = 1.\frac{1}{2^2} + 4.\frac{1}{2^3} + 9.\frac{1}{2^4} + 16.\frac{1}{2^5} + \ldots\ldots\ldots$$

$$\therefore \quad S\left(1 - \frac{1}{2}\right) = 1.\frac{1}{2} + 3.\frac{1}{2^2} + 5.\frac{1}{2^3} + 7.\frac{1}{2^4} + 9.\frac{1}{2^5} + \ldots\ldots$$

$$\frac{S}{2} = 1.\frac{1}{2} + 3.\frac{1}{2^2} + 5.\frac{1}{2^3} + 7.\frac{1}{2^4} + 9.\frac{1}{2^5} + \ldots\ldots$$

$$\frac{S}{4} = \frac{S}{2^2} = 1.\frac{1}{2^2} + 3.\frac{1}{2^3} + 5.\frac{1}{2^4} + 7.\frac{1}{2^5}$$

Subtracting
$$S\left(\frac{1}{2} - \frac{1}{4}\right) = 1.\frac{1}{2} + 2.\frac{1}{2^2} + 2.\frac{1}{2^3} + 2.\frac{1}{2^4} + \ldots\ldots$$

$$= \frac{1}{2} + 2.\frac{1}{2^2}\left\{1 + \frac{1}{2} + \frac{1}{2^2} + \ldots\ldots\right\}$$

$$= 2 + \frac{1}{2}.\frac{1}{1 - \frac{1}{2}} = 3$$

$$S = 12$$

$$\therefore \quad \text{Var}(x) = E(x^2) - \bar{x}^2$$

$$= 12 - 4 = 8$$

$$M_0(t) = E\{e^{tx}\}$$

$$= \sum_{x=1}^{\infty} e^{tx}.\frac{1}{2^x}$$

$$= \sum_{1}^{\infty}\left(\frac{e^t}{2}\right)^x = \frac{e^t/2}{1 - e^t/2}.$$

Ex. 8-10. *An urn contains*
then a fair coin is tossed the n
the expected number of heads.

Sol. Let B_1, B_2, B_3 be the

Let x denote the number
Possible values of x are :

$P($

Now $P(x = 1/B_1) =$ prol

$P(x = 1$

$P(x = 1$

$\therefore$ $P($

$P(x = 2$

$P(x = 2$

$P(x = 2$

$\therefore$ $P($

ıote the number of tosses. Find

$\frac{1}{3}+4^2.\frac{1}{2^4}+\dots\dots\dots$

$.6.\frac{1}{2^4}+25.\frac{1}{2^5}+\dots\dots\dots$

$+16.\frac{1}{2^5}+\dots\dots\dots$

$7.\frac{1}{2^4}+9.\frac{1}{2^5}+\dots\dots$

$7.\frac{1}{2^4}+9.\frac{1}{2^5}+\dots\dots$

$\cdot 7.\frac{1}{2^5}$

$2.\frac{1}{2^4}+\dots\dots$

$\cdot+\dots\dots\}$

Ex. 8-10. *An urn contains balls numbered 1, 2, 3. First a ball is drawn from the urn and then a fair coin is tossed the number of times as the number shown on the drawn ball. Find the expected number of heads.*

Sol. Let B_1, B_2, B_3 be the events that balls numbered 1, 2, 3 are drawn respectively.

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$

Let x denote the number of heads.
Possible values of x are : 0, 1, 2, 3.

$$P(x=1) = P\left(\bigcup_{i=1}^{3}(B_i \cap x=1)\right)$$

$$= \sum_{i=1}^{3} P(B_i \cap x=1)$$

$$= \sum_{i=1}^{3} P(B_i)P(x=1/B_i)$$

Now $P(x=1/B_1) =$ prob. of getting head when coin is tossed once

$$= \frac{1}{2}$$

$P(x=1/B_2) =$ prob. of one head when coin is tossed twice

$$= {}^2c_1\left(\frac{1}{2}\right)^2 = \frac{1}{2}$$

$P(x=1/B_3) =$ prob. of one head when coin is tossed three times

$$= {}^3c_1.\left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$\therefore \quad P(x=1) = \frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{3}{8}$$

$$= \frac{1}{24}\{4+4+3\} = \frac{11}{24}$$

$$P(x=2/B_1) = 0$$

$$P(x=2/B_2) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

$$P(x=2/B_3) = {}^3c_2\left(\frac{1}{2}\right)^3 = 3/8$$

$$\therefore \quad P(x=2) = \sum_{i=1}^{3} P(x=2 \cap B_i)$$

$$= \sum_{i=1}^{3} P(B_i)\, P(x = 2 / B_i)$$

$$= \frac{1}{3}\left\{0 + \frac{1}{4} + \frac{3}{8}\right\} = \frac{5}{24}$$

$$P(x = 3 / B_1) = 0 = P(x = 3 / B_2)$$

$$P(x = 3 / B_3) = \frac{1}{8}$$

$$\therefore \quad P(x = 3) = \frac{1}{3} \times \frac{1}{8} = \frac{1}{24}$$

$$\therefore \quad E(x) = 0.P(x=0) + 1.P(x=1) + 2.P(x=2) + 3.P(x=3)$$

$$= 1.\frac{11}{24} + 2.\frac{5}{24} + 3.\frac{1}{24}$$

$$= 1.$$

Ex. 8-11. *If x has a distribution given by*

$$P(x=0) = P(x=2) = p \text{ and } P(x=1) = 1 - 2p$$

for $0 < p < \frac{1}{2}$, *for what value of p, variance of x is maximum.*

Sol.

$$\bar{x} = E(x) = 0.p + 1.(1-2p) + 2.p$$

$$= 1$$

$$E(x^2) = 0^2.p + 1^2.(1-2p) + 2^2.p$$

$$= 1 - 2p + 4p = 1 + 2p$$

$$\therefore \quad \mu_2 = \text{var}(x) = E(x^2) - \bar{x}^2$$

$$= 1 + 2p - 1$$

$$= 2p.$$

It will be maximum at $p = \frac{1}{2}$.

Ex. 8-12. *Consider an experiment of rolling of two six faced die. Let x denote the absolute difference of the upturned faces. Find the density function of x. Also find $E(x)$.*

Sol. Possible values of x are :

0, 1, 2, 3, 4, 5.

For $x = 0$, different possibilities are

(1, 1); (2, 2); (3, 3); (4, 4); (5, 5), (6, 6)

$\therefore$ prob. for $x = 0$ is $\frac{6}{36}$.

For $x = 1$, possibilities are :

(1, 2); (2, 3); (3, 4); (4, 5); (5, 6)

(2, 1), (3, 2), (4, 3), (5, 6), (6, 5)

$\therefore$ prob. for $x = 1$ is $\frac{10}{36}$.

Similarly other probabilities aı

$\therefore \quad E(x)$

Ex. 8-13. *A coin is tossed four ti immediately by a tail. Find distribu*

Sol. Possible values of x are : 0

For $x = 0$: Possibilities are

HHHH; THHH; TTH,

$\therefore \quad P(x = 0)$

For $x = 1$: Different possibilitie:

(HT) HH; (HT) TH;

H(HT)H; H(HT)T; T

HH(HT); TH(HT); T

$P(x = 1)$

For $x = 2$: Only possibility is

(HT) HT

$\therefore \quad P(x = 2)$

$\therefore \quad E(x)$

$E(x^2)$

$\therefore \quad \text{var}(x)$

Ex. 8-14. *A and B throw with o player who first throws 6. If A has th*

Sol. *A* can win in 1st, 3rd, 5th,....

$\frac{1}{6}, \left(\frac{5}{6}\right)^2 \frac{1}{6}, \left(\frac{5}{6}\right)^4 \frac{1}{6}, \ldots\ldots$

$\therefore$ *A*'s chance of success

$/ B_i)$

$= 1) + 2.P(x = 2) + 3.P(x = 3)$

$) = 1 - 2p$

um.

p

$2^2.p$

ix faced die. Let x denote the unction of x. Also find $E(x)$.

(6, 6)

5)

5)

$\therefore$ prob. for $x = 1$ is $\frac{10}{36}$.

Similarly other probabilities are :

$$\frac{8}{36}, \frac{6}{36}, \frac{4}{36}, \frac{2}{36}$$

$$\therefore \quad E(x) = 0.\frac{6}{36} + 1.\frac{10}{36} + 2.\frac{8}{36} + 3.\frac{6}{36} + 4.\frac{4}{36} + 5.\frac{2}{36}$$

$$= \frac{70}{36}.$$

Ex. 8-13. *A coin is tossed four times. Let x denote the number of times a head is followed immediately by a tail. Find distribution, mean and variance of x.*

Sol. Possible values of x are : 0, 1, 2

For $x = 0$: Possibilities are

HHHH; THHH; TTHH; TTTH, TTTT,

$$\therefore \quad P(x = 0) = \frac{5}{16}$$

For $x = 1$: Different possibilities are :

(HT) HH; (HT) TH; (HT) TT;
H(HT)H; H(HT)T; T(HT)H; T(HT)T
HH(HT); TH(HT); TT(HT)

$$P(x = 1) = \frac{10}{16}$$

For $x = 2$: Only possibility is

(HT) HT

$$\therefore \quad P(x = 2) = 1/16$$

$$\therefore \quad E(x) = 0.\frac{5}{16} + 1.\frac{10}{16} + 2.\frac{1}{16} = \frac{3}{4}$$

$$E(x^2) = 0^2.\frac{5}{16} + 1^2.\frac{10}{16} + 2^2.\frac{1}{16} = \frac{7}{8}$$

$$\therefore \quad \text{var}(x) = E(x^2) - [E(x)]^2$$

$$= \frac{7}{8} - \frac{9}{16} = \frac{5}{16}.$$

Ex. 8-14. *A and B throw with one die for a prize of Rs. 11 which is to be won by the player who first throws 6. If A has the first throw, what are their respective expectations ?*

Sol. *A* can win in 1st, 3rd, 5th,..........trials with respective chances

$$\frac{1}{6}, \left(\frac{5}{6}\right)^2 \frac{1}{6}, \left(\frac{5}{6}\right)^4 \frac{1}{6}, \ldots\ldots$$

$\therefore$ *A*'s chance of success

$$= \frac{1}{6}\left\{1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \ldots\ldots\right\}$$

$$= \frac{1}{6}\left\{\frac{1}{1-\frac{25}{36}}\right\} = \frac{6}{11}$$

Since there are only two players, and total probability is unity. B's chance of success

$$= 1 - \frac{6}{11} = \frac{5}{11}$$

$\therefore$ A's expectation $= \frac{6}{11} \times 11 = \text{Rs. } 6$

and B's expectation $= \frac{5}{11} \times 11 = \text{Rs. } 5.$

Ex. 8-15. *If x is a random variate which assumes values 1, 2, 3, 4,........ with respective probabilities given by*

$$P(x = k) = q^{k-1}p, \; q + p = 1,$$

find $E(x)$

Sol.

$$E(x) = \sum_{k=1}^{\infty} k.q^{k-1}.p$$

$$= p\sum_{k=1}^{\infty} kq^{k-1}$$

$$= p\{1 + 2q + 3q^2 + ...\}$$

$$= p(1-q)^{-2} = \frac{1}{p}.$$

Ex. 8-16. *Let a random variate x take the values*

$$x_k = (-1)^k \frac{2^k}{k}, \; k = 1, 2, 3, \ldots\ldots\ldots\ldots$$

with probabilities $p_k = 2^{-k}$. *Find E(x).*

Sol. Total prob. $= \sum_{k=1}^{\infty} 2^{-k}$

$$= \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots\ldots$$

$$= \frac{1/2}{1-1/2} = 1$$

$$\therefore \quad E(x) = \sum_{k=1}^{\infty} (-1)^k \frac{2^k}{k} . 2^{-k}$$

Ex. 8-17. *A random variable*

proportional to $\frac{1}{3^n}$. *Find* $E(x)$.

Sol. Let $P(x = n) = \frac{\lambda}{3^n}$,

Where λ is constant of propc

$\therefore$ Total pr

Since total prob. $= 1$, λ is giv

$\Rightarrow$

$\therefore$ $P(x =$

$\therefore$ $E($

lity is unity. B's chance of success

alues 1, 2, 3, 4,........ with respective

$= 1,$

$+\ldots\}$

....

$+\ldots\ldots$

$^{-k}$

$$= \sum_{k=1}^{\infty} \frac{(-1)^k}{k}$$

$$= -1+\frac{1}{2}-\frac{1}{3}+\frac{1}{4}\ldots.$$

$$= -\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots\right)$$

$$= -\log 2.$$

Ex. 8-17. *A random variable x can assume any positive integral value n with a probability proportional to* $\frac{1}{3^n}\cdot$ *Find* $E(x)$.

Sol. Let $P(x=n) = \frac{\lambda}{3^n}$,

Where λ is constant of proportionality.

$$\therefore \qquad \text{Total prob.} = \lambda \sum_{n=1}^{\infty} \frac{1}{3^n}$$

$$= \lambda\left\{\frac{1}{3}+\frac{1}{3^2}+\frac{1}{3^3}+\ldots.\right\}$$

$$= \lambda \frac{1/3}{1-1/3} = \frac{\lambda}{2}$$

Since total prob. = 1, λ is given by

$$\frac{\lambda}{2} = 1$$

$$\Rightarrow \lambda = 2.$$

$$\therefore \qquad P(x=n) = \frac{2}{3^n}$$

$$\therefore \qquad E(x) = 2\sum_{n=1}^{\infty} n.\frac{1}{3^n}$$

$$= 2\left\{\frac{1}{3}+2.\frac{1}{3^2}+3.\frac{1}{3^3}+\ldots..\right\}$$

$$= \frac{2}{3}\left\{1+2.\frac{1}{3}+3.\frac{1}{3^2}+\ldots.\right\}$$

$$= \frac{2}{3}\left(1-\frac{1}{3}\right)^{-2} = \frac{2/3}{(2/3)^2} = \frac{3}{2}.$$

8.3. Laws of Expectation. *Two basic laws of expectation are*

(1) $E(x+y)=E(x)+E(y)$

(2) $E(xy)=E(x)E(y)$

provided x and y are independent.

Proof. (1) Let x and y be two stochastic variates with probability distributions

$$\begin{matrix} x \rightarrow \\ p \rightarrow \end{matrix}\begin{pmatrix} x_1 x_2 \ldots\ldots x_m \\ p_1 p_2 \ldots\ldots p_m \end{pmatrix} \quad \textit{and} \quad \begin{matrix} y \rightarrow \\ p \rightarrow \end{matrix}\begin{pmatrix} y_1 y_2 \ldots\ldots\ldots\ldots y_n \\ P_1 P_1 \ldots\ldots\ldots\ldots P_n \end{pmatrix}$$

Let $\qquad z = x + y$

Then z will also be a stochastic variate.

Let $\qquad z_{ij} = x_i y_j$

Let p_{ij} be the probability of z taking a value z_{ij}.

Let A_i be the event that x takes the value x_i and A_{ij} be the event that z takes the value z_{ij}.

Then $(A_i) = (A_{i1} + A_{i2} + \ldots\ldots + A_{in})$

$\therefore \; P(A_i) = P(A_{i1} + A_{i2} + \ldots\ldots + A_{in})$

Since z can take only one value at a time the events $A_{i1}, A_{i2}, \ldots\ldots A_{in}$ are mutually exclusive.

$\therefore$ By total probability theorem

$$P(A_i) = P(A_{i1}) + P(A_{i2}) + \ldots\ldots + P(A_{in})$$

i.e.,
$$p_i = p_{i1} + p_{i2} + \ldots\ldots + p_{in}$$
$$i = 1, 2, \ldots\ldots m$$

Similarly
$$P_j = p_{1j} + p_{2j} + \ldots + p_{mj} \qquad j = 1, 2, \ldots\ldots\ldots\ldots\ldots n$$

Now
$$E(x+y) = E(z) = \sum_{i=1}^{in} \sum_{j=1}^{n} z_{ij} p_{ij}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i + y_j) p_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i p_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n} y_j p_{ij}$$

$$= \sum_{i=1}^{m} x_i \{p_{i1} + p_{i2} + \ldots\ldots + p_{in}\} + \sum_{j=1}^{n} y_j \{p_{1j} + p_{2j} + \ldots\ldots + p_{mj}\}$$

$$= \sum_{i=1}^{m} x_i p_i + \sum_{j=1}^{n} y_j P_j$$

$$= E(x) + E(y).$$

(2) Def. Two stochastic variates are said to be independent, if the probability of either taking a particular value does not depend on what value the other variate takes.

Let x and y be two stochastic variates with probability distributions.

$\begin{matrix} x \rightarrow \\ p \rightarrow \end{matrix}\Big(\begin{matrix} x_1 x \\ p_1 p \end{matrix}$

Let

Then z will also be a stocha

Let

Let p_{ij} be the probability o

by compound probability theore

p_{ij} =

$\therefore \qquad E(xy) = E(z)$ =

=

=

=

Remark : These laws can b

$x_1, x_2 \ldots x_n$ be n chance variates,

$E(x_1 + x_2 + + x_n)$ =

and $\qquad E(x_1 x_2 \ldots x_n)$ =

Ex. 8-18. *Find expected valu*

Sol. Let x_i be the number c

dice = $x_1 x_2 \ldots\ldots x_n$.

Therefore, expected value o:

= prodı

This because $x_1, x_2 \ldots\ldots x_n$

independent of the number obtai

But expected value of x_i =

Therefore, expected value o

=

Ex. 8-19. *Find the mathemat*

dice together.

Sol. Let x_i be the number o

will be

ion are

›robability distributions

$$\begin{matrix} ....y_n \\ ....P_n \end{matrix}\Big)$$

$_{ij}$ be the event that z takes the

s $A_{i1}, A_{i2}, ...A_{in}$ are mutually

$A_{in})$

j = 1, 2,................n

$$\sum_{j=1}^{n} x_i p_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n} y_j p_{ij}$$

$$\} + \sum_{j=1}^{n} y_j \{p_{1j} + p_{2j} + ...+ p_{mj}\}$$

ıdent, if the probability of either
e other variate takes.
distributions.

$$\begin{matrix} x \to \\ p \to \end{matrix}\begin{pmatrix} x_1 x_2 ....x_m \\ p_1 p_2 ....p_m \end{pmatrix} \text{and} \begin{matrix} y \to \\ p \to \end{matrix}\begin{pmatrix} y_1 y_2 .....y_n \\ P_1 P_2 ......P_n \end{pmatrix}$$

Let $$z = xy$$

Then z will also be a stochastic variate.

Let $$z_{ij} = x_i y_j$$

Let p_{ij} be the probability of z taking a value z_{ij}. Then since x and y are independent, by compound probability theorem.

$$p_{ij} = p_i P_j$$

$$\therefore \qquad E(xy) = E(z) = \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} p_{ij}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i y_j p_i P_j)$$

$$= \left(\sum_{i=1}^{m} x_i p_i\right)\left(\sum_{j=1}^{n} y_j P_j\right)$$

$$= E(x)E(y).$$

Remark : These laws can be generalized to any finite number of variates namely: if $x_1, x_2 ... x_n$ be n chance variates, then

$$E(x_1 + x_2 + + x_n) = E(x_1) + E(x_2) + ... + E(x_n)$$

and $$E(x_1 x_2 ... x_n) = E(x_1) E(x_2) ... E(x_n) \text{ provided } x\text{'s are independent.}$$

Ex. 8-18. *Find expected value of the product of points obtained on rolling n dice together.*

Sol. Let x_i be the number of points obtained on ith die. Then product of points on n dice $= x_1 x_2 ......x_n$.

Therefore, expected value of the product of points obtained

= product of the expected values of x_i

This because $x_1, x_2 ......x_n$ are independent as number obtained on one die is independent of the number obtained on other dice

But expected value of $x_i = \dfrac{7}{2}$ (See Ex. 8-2)

Therefore, expected value of the product of points obtained

$$= \left(\frac{7}{2}\right)^n.$$

Ex. 8-19. *Find the mathematical expectation of the sum of points obtained on rolling n dice together.*

Sol. Let x_i be the number of points obtained on ith die. Then sum of points on n dice will be

$$s = x_1 + x_2 + \ldots\ldots\ldots\ldots + x_n$$

Therefore, expected value of s

$$= \text{sum of the expected values of } x_1, x_2 \ldots\ldots x_n$$

Now expected value of $x_i = \dfrac{7}{2}$ (See Ex. 8-2)

Therefore, expected value of s

$$= \left(\frac{7}{2}\right)n$$

$$= \frac{7n}{2}.$$

Ex. 8-20. *If p_i be the probability of success for ith trial, find the expectation of the number of successes in n independent trials.*

Sol. Associate with every trial a variable which has the value '1' in case of success and the value '0' in case of failure. If $x_1, x_2 \ldots\ldots x_n$ be the variables attached to trials 1, 2,...n, the number of successes in n trials is given by

$$m = x_1 + x_2 + \ldots + x_n$$

$$\therefore \quad E(m) = E(x_1) + E(x_2) + \ldots\ldots + E(x_n)$$

Since x_i can take only two values '1' and '0' with respective probabilities p_i and $1 - p_i$ its expectation is given by

$$E(x_i) = 1.p_i + 0.(1 - p_i)$$

$$= p_i$$

$$\therefore \quad E(m) = p_1 + p_2 + \ldots\ldots\ldots + p_n.$$

Ex. 8-21. *Find the expectation of the number of white balls among c balls drawn from an urn containing a white and b black balls.*

Sol. Associate with every ball a variable which has the value '1' if it is white and the value '0' otherwise. If $x_1, x_2 \ldots\ldots x_c$ be the variables attached to c balls drawn, the number of white balls is given by

$$m = x_1 + x_2 + \ldots\ldots + x_c$$

$$\therefore \quad E(m) = E(x_1) + E(x_2) + \ldots\ldots + E(x_c)$$

Now the probability that the ith ball drawn will be white when nothing is known of the other balls

$$= \frac{a}{a+b}$$

$$\therefore \quad E(x_i) = 1\frac{a}{a+b} + 0.\left\{1 - \frac{a}{a+b}\right\}$$

$$= \frac{a}{a+b} \text{ for all } i$$

$$\therefore \quad E(m) = \frac{ca}{a+}$$

Ex. 8-22. *A box contains 2^n ti*
$i = 0, 1, \ldots\ldots n$.

A group of m tickets is drawn at r
of numbers on them.

Sol. Let $x_1, x_2, \ldots\ldots x_m$ be the nu

Consider x_k.

Its possible values are 0, 1, 2,.....

Since there are ${}^nc_i$ tickets bearii

$$\frac{{}^nc_{(}}{2^n}$$

$$\therefore \quad E(x_k) :$$

$$\therefore \quad E(x_1 + x_2 + \ldots\ldots + x_m) :$$

(ii) $E\left(x_k^2\right) :$

$_2 \ldots\ldots x_n$

(See Ex. 8-2)

al, find the expectation of the

alue '1' in case of success and
s attached to trials 1, 2,...n, the

spective probabilities p_i and

alls among c balls drawn from

value '1' if it is white and the
d to c balls drawn, the number

when nothing is known of the

$$\therefore \qquad E(m) = \frac{ca}{a+b}.$$

Ex. 8-22. *A box contains 2^n tickets among which ${}^n c_i$ tickets bear the number i, $i = 0, 1, \ldots\ldots n$.*

A group of m tickets is drawn at random. Find the expectation and variance of the sum of numbers on them.

Sol. Let $x_1, x_2, \ldots\ldots x_m$ be the numbers on m tickets drawn.

Consider x_k.

Its possible values are 0, 1, 2,.....n.

Since there are ${}^n c_i$ tickets bearing number i, probabilities of these values are

$$\frac{{}^n c_0}{2^n}, \frac{{}^n c_1}{2^n}, \frac{{}^n c_2}{2^n} \ldots \frac{{}^n c_n}{2^n}$$

$$\therefore \qquad E(x_k) = \frac{1}{2^n}\left\{0^n c_0 + 1^n c_1 + 2^n c_2 + \ldots\ldots + n^n c_n\right\}$$

$$= \frac{1}{2_n}\left\{1.\, n + 2.\frac{n(n-1)}{2!} + \ldots\ldots + n\right\}$$

$$= \frac{n}{2^n}\left\{1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \ldots\ldots + 1\right\}$$

$$= \frac{n}{2^n}(1+1)^{n-1} = \frac{n}{2}$$

$$\therefore \qquad E(x_1 + x_2 + \ldots\ldots + x_m) = E(x_1) + \ldots\ldots + E(x_m)$$

$$= \frac{mn}{2}$$

(ii) $$E\left(x_k^2\right) = \sum_{i=0}^{n} i^2 \frac{{}^n c_i}{2^n}$$

$$= \sum_{i=0}^{n}\{i(i-1)+i\}\frac{{}^n c_i}{2^n}$$

$$= \frac{1}{2^n}\sum_{i=0}^{n} i(i-1)\,{}^n c_i + E(x_k)$$

$$= \frac{1}{2^n}\{2.1.{}^n c_2 + 3.2.{}^n c_3 + \ldots\ldots + n(n-1).{}^n c_n\} + \frac{n}{2}$$

$$= \frac{n(n-1)}{2^n}\{1 + (n-2) + \ldots\ldots + 1\} + \frac{n}{2}$$

$$= \frac{n(n-1)}{2^n}(1+1)^{n-2} + \frac{n}{2}$$

$$= \frac{n(n-1)}{4} + \frac{n}{2} = \frac{n(n+1)}{4}$$

To find cov. (x_k, x_l), simultaneous values of x_k and x_l are to be considered.

Here prob. of x_k taking value i and x_l taking a value j is

$$\frac{{}^n c_i \, . \, {}^n c_j}{2^n(2^n-1)} \quad \text{if } i \neq j.$$

and
$$\frac{{}^n c_i \, ({}^n c_i - 1)}{2^n(2^n-1)} \quad \text{if } i = j$$

$$\therefore \quad E(x_k x_l) = \sum_{i=0}^{n} \sum_{\substack{j=0 \\ j \neq i}}^{\bar{n}} i.j \frac{{}^n c_i \; {}^n c_j}{2^n\,(2^n-1)} + \sum_{i=0}^{n} i^2 \frac{{}^n c_i . ({}^n c_i - 1)}{2^n(2^n-1)}$$

$$= \frac{1}{2^n(2^n-1)} \left[\sum_{i=0}^{n} i.{}^n c_i \left\{ \sum_{j=0}^{n} j.{}^n c_j - i.{}^n c_i \right\} + \sum_{i=0}^{n} i^2 . {}^n c_i ({}^n c_i - 1) \right]$$

$$= \frac{1}{2^n(2^n-1)} \left[\sum_{i=0}^{n} i.{}^n c_i \frac{n}{2} . 2^n - \sum_{i=0}^{n} i^2 . {}^n c_i \right]$$

$$= \frac{1}{2^n - 1} \left[\frac{n^2}{4} . 2^n - \frac{n(n+1)}{4} \right]$$

Let
$$S = x_1 + \ldots\ldots . x_m$$

$$\therefore \quad E(S^2) = E\left(\sum_{i=1}^{m} x_i \right)^2$$

$$= E\left\{ \sum_{k=1}^{m} x_k{}^2 + \sum_{k} \sum_{\substack{l \\ k \neq l}} x_k x_l \right\}$$

$$= \sum_{k=1}^{m} \frac{n(n+1)}{4} + \sum_{k} \sum_{\substack{l \\ k \neq l}} \frac{1}{2^n - 1} \left\{ \frac{n^2}{4} . 2^n - \frac{n(n+1)}{4} \right\}$$

$$= \frac{mn(n+1)}{4} + \frac{m(m-1)}{2^n - 1} . \frac{n}{4} \{ n.2^n - n - 1 \}$$

$$= \frac{mn(n+1)}{4} + m(m-1) . \frac{n^2}{4} - \frac{m(m-1)n}{4(2^n - 1)}$$

$= \frac{mn}{4} \{$

$= \frac{mn(}{}$

$\therefore \quad \text{Var } (S) = E(S^2$

$= \frac{mn(}{}$

$= \frac{mn}{4} -$

Ex. 8-23. *Balls are taken on*
until the first white ball is draw
preceding the first white ball is

Sol. Let x be the number c
values of x are 0, 1, 2,....b
Prob. of x taking the value '
= prob. of drawing a white b

Prob. of x taking the value '
= Prob. of drawing a black b

in se

Prob. of x taking the value '
= prob. of black balls in firs

third dr

and so on. In general, prob.

$= \frac{b}{a+b}$

$\therefore$ Expected value of x

$= 0.\frac{a}{a+}$

+.....

$= \frac{ab}{a+b}$

$$= \frac{mn}{4}\{n+1+n(m-1)\} - \frac{nm(m-1)}{4(2^n-1)}$$

$$= \frac{mn(mn+1)}{4} - \frac{nm(m-1)}{4(2^n-1)}$$

$$\therefore \quad \text{Var }(S) = E(S^2) - \{E(S)\}^2$$

$$= \frac{mn(mn+1)}{4} - \frac{nm(m-1)}{4(2^n-1)} - \left(\frac{mn}{2}\right)^2$$

$$= \frac{mn}{4} - \frac{nm(m-1)}{4(2^n-1)}.$$

Ex. 8-23. *Balls are taken one by one out of an urn containing a white and b black balls until the first white ball is drawn. Show that the expectation of the number of black balls preceding the first white ball is* $\frac{b}{a+1}$.

Sol. Let x be the number of black balls drawn before first white ball. The possible values of x are 0, 1, 2,....b

Prob. of x taking the value '0'

$= $ prob. of drawing a white ball in first draw $= \frac{a}{a+b}$

Prob. of x taking the value '1'

$=$ Prob. of drawing a black ball in first draw and a white ball

in second draw $= \frac{b}{a+b} \cdot \frac{a}{a+b-1}$

Prob. of x taking the value '2'

$=$ prob. of black balls in first two draws and a white ball in

third draw $= \frac{b}{a+b} \cdot \frac{b-1}{a+b-1} \cdot \frac{a}{a+b-2}$

and so on. In general, prob. of x taking the value 'r'

$$= \frac{b}{a+b} \cdot \frac{b-1}{a+b-1} \cdots\cdots \frac{b-\overline{r-1}}{a+b-\overline{r-1}} \cdot \frac{a}{a+b-r}$$

$\therefore$ Expected value of x

$$= 0.\frac{a}{a+b} + 1.\frac{b}{a+b} \cdot \frac{a}{a+b-1} + 2.\frac{b(b-1)}{(a+b)(a+b-1)} \cdot \frac{a}{(a+b-2)}$$

$$+ \ldots\ldots r\frac{b(b-1)\ldots\ldots(b-\overline{r-1})}{(a+b)(a+b-1)\ldots\ldots(a+b-\overline{r-1})}\frac{a}{a+b-r} + \ldots$$

$$= \frac{ab}{a+b}\left[\frac{1}{a+b-1} + 2\frac{(b-1)}{(a+b-1)(a+b-2)} + \ldots\ldots\right.$$

$$+r\frac{(b-1)(b-2).....(b-\overline{r-1})}{(a+b-1)(a+b-2)......(a+b-r)}+.......\Bigg]$$

Let $$U_r = \frac{r(b-1)(b-2).....(b-\overline{r-1})}{(a+b-1)(a+b-2).....(a+b-r)}$$

$$= \frac{[A(r-1)+B](b-1)(b-2).....(b-\overline{r-1})}{(a+b-1)(a+b-2).......(a+b-\overline{r-1})}$$

$$-\frac{[Ar+B](b-1)(b-2).....(b-r)}{(a+b-1)(a+b-2).....(a+b-r)}$$

$\therefore$ $$r = [A(r-1)+B](a+b-r)-[Ar+B](b-r)$$

Equating co-efficients of r

$$1 = A(a+b)+A-A.b$$

or $$A = \frac{1}{a+1}$$

Equating terms independent of r

$$0 = (B-A)(a+b)-Bb$$

or $$B = \frac{a+b}{a}A$$

$\therefore$ $$U_r = \frac{A\left[\overline{r-1}+\frac{a+b}{a}\right](b-1)(b-2)....(b-\overline{r-1})}{(a+b-1)(a+b-2)....(a+b-\overline{r-1})}$$

$$-\frac{A\left[r+\frac{a+b}{a}\right](b-1)(b-2)....(b-r)}{(a+b-1)(a+b-2)....(a+b-r)}$$

$\therefore$ $U_2+U_3+.......U_b$

$$= \frac{A\left[1+\frac{a+b}{a}\right](b-1)}{(a+b-1)} = \frac{\left[1+\frac{a+b}{a}\right](b-1)}{(a+1)(a+b-1)}$$

$\therefore$ Expected value of x

$$= \frac{ab}{a+b}\left[\frac{1}{a+b-1}+\frac{\left[1+\frac{a+b}{a}\right](b-1)}{(a+1)(a+b-1)}\right]$$

$$= \frac{ab}{a+b}\left[\frac{a+1+b-1+\frac{(a+b)(b-1)}{a}}{(a+1)(a+b-1)}\right]$$

Ex. 8-24. *A bag contains a coi*
value is m. A person draws one a
expectations.

Sol. Let the total number of c
Then the average value of coi

Now prob. of drawing a coin

Since in first draw any coin m

If the coin M does not appear
$\therefore$ Chance of second draw = (

Since there will be $(n-1)$ coi
draw, when it is known that secon

$\therefore$ By compound prob. theore

$\therefore$ Expectation from second d

Similarly expectation from th

$$= \frac{b}{a+1}.$$

Ex. 8-24. *A bag contains a coin of value M and a number of other coins whose aggregate value is m. A person draws one at a time till he draws the coin M. Find the value of his expectations.*

Sol. Let the total number of coins in a bag = n.

Then the average value of coins other than that of value M

$$= \frac{m}{n-1}$$

Now prob. of drawing a coin in first draw

$$= \frac{1}{n}$$

Since in first draw any coin may appear, expectation from first draw

$$= \left\{M + \frac{m}{n-1} + \frac{m}{n-1} + \ldots + (n-1) \text{ times}\right\} \cdot \frac{1}{n}$$

$$= \frac{M+m}{n}.$$

If the coin M does not appear in first draw, second draw is to be made.

$\therefore$ Chance of second draw = Chance of not drawing the coin M in first draw

$$= 1 - \frac{1}{n} = \frac{n-1}{n}$$

Since there will be $(n-1)$ coins before second draw, chance of drawing a coin in second draw, when it is known that second draw is to be made

$$= \frac{1}{n-1}$$

$\therefore$ By compound prob. theorem, prob. of drawing a coin in 2nd draw

$$= \frac{n-1}{n} \cdot \frac{1}{n-1}$$

$$= \frac{1}{n}$$

$\therefore$ Expectation from second draw

$$= \frac{1}{n}\left\{M + \frac{m}{n-1} + \ldots\ldots + (n-2) \text{ times}\right\}$$

$$= \frac{1}{n}\left\{M + \frac{(n-2)m}{n-1}\right\}.$$

Similarly expectation from third draw

$$= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \frac{1}{n-2}\left\{M + \frac{n-3}{n-1}m\right\}$$

$$= \frac{1}{n}\left\{M + \frac{n-3}{n-1}m\right\}$$

and so on.

Finally expectation from last draw

$$= \frac{M}{n}$$

$\therefore$ Total expectation

$$= \frac{1}{n}\left[(M+m) + \left\{M + \frac{n-2}{n-1}m\right\} + \left\{M + \frac{n-3}{n-1}m\right\} + \ldots + M\right]$$

$$= M + \frac{1}{n}.m\left\{1 + \frac{n-2}{n-1} + \frac{n-3}{n-1} + \ldots + \frac{1}{n-1}\right\}$$

$$= M + \frac{1}{n(n-1)}m\{(n-1)+(n-2)+\ldots+1\}$$

$$= M + \frac{1}{n(n-1)}m.\frac{n(n-1)}{2}$$

$$= M + \frac{m}{2}.$$

Ex. 8-25. *Show that* $E|x| \geq |E(x)|$

Sol. $|E(x)| = |\Sigma px|$

$$\leq \Sigma|px| = \Sigma p|x| = E|x|.$$

Ex. 8-26. *Show by an example that the mathematical expectation need not be finite.*

Sol. Consider the prob. dist.

$$P(x) = \frac{e^{-1}}{x!}, \qquad x = 0, 1, 2, \ldots\ldots\ldots\ldots\infty$$

Here $$E(x!) = \sum_{x=0}^{\infty} \frac{e^{-1}}{x!}.x! = e^{-1}\sum_{x=0}^{\infty} 1$$

which is not finite.

Ex. 8-27. *Show that* $E(x^2) \geq \{E(x)\}^2$.

Sol. We have

$$E(x-\bar{x})^2 = \Sigma p(x-\bar{x})^2 \geq 0$$

i.e., $E(x^2 + \bar{x}^2 - 2x\bar{x}) \geq 0.$

i.e., $E(x^2) + \bar{x}^2 - 2\bar{x}E(x) \geq 0$

i.e., $E(x^2) + \bar{x}^2 - 2\bar{x}^2 \geq 0$

$\therefore$ $E(x^2) \geq \bar{x}^2 = \{E(x)\}^2$

Ex. 8-28. *For any variates x and y show that*

$$\left\{E(x+y)^2\right\}^{\frac{1}{2}} \leq \left\{E(x^2)\right\}^{\frac{1}{2}} + \left\{E(y^2)\right\}^{\frac{1}{2}}$$

Sol. We have $(ax-y)^2 \geq 0,$

$\Rightarrow$ $E(ax-y)^2 \geq 0$ $(\because$

$\Rightarrow$ $a^2E(x^2) + E(y^2) - 2a$

Put

$\therefore$ $\frac{\{E(xy)\}^2}{\{E(x^2)\}^2}.E($

$\Rightarrow$ $\{E(xy)\}$

$\Rightarrow$ $E(xy) \leq$

Now $E(x+y$

$\Rightarrow$ $\{E(x+y)^2$

Ex. 8-29. *For independent*

if and only if

E

Sol. If E

$\therefore$ Var (

Conversely, let

Var $(xy) = $ Var (x) Var (y)

$\Rightarrow$ $E(x^2y^2) - \bar{x}^2\bar{y}^2 = \{E($

$\Rightarrow$ $E(x^2)E(y^2) - \bar{x}^2\bar{y}^2 =$

$\Rightarrow$ $\bar{x}^2E(y^2) + \bar{y}^2E(x^2) -$

$\Rightarrow$ $\bar{x}^2\{E(y^2) - \bar{y}^2\} + \bar{y}^2\{$

$\Rightarrow$ $\bar{x}^2$ Var $(y) + \bar{y}^2$ Var (

$+\left\{M+\frac{n-3}{n-1}m\right\}+....+M\Bigg]$

$...+\frac{1}{n-1}\Bigg\}$

$-1)+(n-2)+....+1\}$

$\frac{n-1)}{2}$

expectation need not be finite.

$\sum_{=0}^{\infty} 1$

$y^2)\}^{\frac{1}{2}}$

Sol. We have $(ax-y)^2 \geq 0$, for all real constants 'a'.

$$\Rightarrow \quad E(ax-y)^2 \geq 0 \qquad (\because \text{ prob. } \geq 0)$$

$$\Rightarrow \quad a^2E(x^2)+E(y^2)-2aE(xy) \geq 0$$

Put $$a = \frac{E(xy)}{E(x^2)}$$

$$\therefore \quad \frac{\{E(xy)\}^2}{\{E(x^2)\}^2}\cdot E(x^2)+E(y^2)-\frac{2\{E(xy)\}^2 \geq 0}{E(x^2)}$$

$$\Rightarrow \quad \{E(xy)\}^2 \leq E(x^2)E(y^2)$$

$$\Rightarrow \quad E(xy) \leq \sqrt{E(x^2)E(y^2)}$$

Now
$$\begin{aligned} E(x+y)^2 &= E(x^2)+E(y^2)+2E(xy) \\ &\leq E(x^2)+E(y^2)+2\sqrt{E(x^2)E(y^2)} \\ &= \left\{\sqrt{E(x^2)}+\sqrt{E(y^2)}\right\}^2 \end{aligned}$$

$$\Rightarrow \quad \left\{E(x+y)^2\right\}^{\frac{1}{2}} \leq \left\{E(x^2)\right\}^{\frac{1}{2}}+\left\{E(y^2)\right\}^{\frac{1}{2}}$$

Ex. 8-29. *For independent and non-degenerate variates x and y, show that*

$$\text{Var}(xy) = \text{Var}(x).\text{Var}(y)$$

if and only if

$$E(x) = 0 = E(y)$$

Sol. If $E(x) = 0 = E(y)$,

$$\begin{aligned} \therefore \quad \text{Var}(xy) &= E\{xy\}^2-\{E(xy)\}^2 \\ &= E(x^2y^2)-\{E(x)E(y)\}^2 \\ &= E(x^2)\,E(y^2) \\ &= \text{Var}(x)\ \text{Var}(y) \end{aligned}$$

Conversely, let

$\text{Var}(xy) = \text{Var}(x)\,\text{Var}(y)$

$$\Rightarrow \quad E(x^2y^2)-\bar{x}^2\bar{y}^2 = \left\{E(x^2)-\bar{x}^2\right\}\left\{E(y^2)-\bar{y}^2\right\}$$

$$\Rightarrow \quad E(x^2)E(y^2)-\bar{x}^2\bar{y}^2 = E(x^2)E(y^2)-\bar{x}^2E(y^2)-\bar{y}^2E(x^2)+\bar{x}^2\bar{y}^2$$

$$\Rightarrow \quad \bar{x}^2E(y^2)+\bar{y}^2E(x^2)-2\bar{x}^2\bar{y}^2 = 0$$

$$\Rightarrow \quad \bar{x}^2\left\{E(y^2)-\bar{y}^2\right\}+\bar{y}^2\left\{E(x^2)-\bar{x}^2\right\} = 0$$

$$\Rightarrow \quad \bar{x}^2\ \text{Var}(y)+\bar{y}^2\ \text{Var}(x) = 0 \qquad ...(1)$$

Since x and y are non-degenerate,

Var $(x) > 0$, Var $(y) > 0$.

$\therefore$ (1) $\Rightarrow$ $\bar{x} = \bar{y} = 0$.

8.4. Moment Generating Function. *The moment generating function (m.g.f.) of the chance variate about the point 'a' is defined to be E* $\left\{e^{t(x-a)}\right\}$ *and is denoted by* $M_a(t)$, *where t is the real parameter.*

Cumulative Function. *The cumulative function about* $x = a$ *is defined by* $K_a(t) = \log M_a(t)$. *If* $K_a(t)$ *can be expanded as a convergent series in powers of t viz.,*

$$K_a(t) \equiv k_1 t + k_2 \frac{t^2}{2!} + k_3 \frac{t^3}{3!} + \ldots$$

the co-efficients k_1, k_2 *etc., are called first cumulent, second cumulent etc., of the dist.*

Ex. 8-30. (*i*) *Show that* $M_a(t) = e^{-at} M_0(t)$

(*ii*) *Discuss the effect of change of origin and scale on M.G.F.*

(*iii*) *Show that the m.g.f. of the sum of 'n' independent variates is the product of their moment generating functions.*

(*iv*) *Show that*

$$M_a(t) = \sum_{r=0}^{\infty} \mu'_r(a) \frac{t^r}{r!}$$

Sol. (*i*)
$$\begin{aligned} M_a(t) &= E\{e^{t(x-a)}\} \\ &= E\{e^{tx} . e^{-at}\} \\ &= e^{-at} E\{e^{tx}\} \\ &= e^{-at} M_0(t) \end{aligned}$$

(*ii*) The transformation corresponding to change of origin and scale is

$$X = \frac{x-a}{h}$$

where 'a' corresponds to change of origin and 'h' to change of scale. Both a and h are constants. X is a new variate to which x transforms.

$$\therefore \qquad x = a + hX$$

$$\begin{aligned} \therefore \qquad M_0(t) \text{ of } x &= E\{e^{tx}\} \\ &= E\{e^{t(a+hX)}\} \\ &= e^{at} E\{e^{(th)X}\} \\ &= e^{at} \{M_0(th) \text{ of } X\} \end{aligned}$$

(*iii*) Let $x_1, x_2, \ldots\ldots x_n$ be n independent chance variates.

Let $$X = x_1 + x_2 + \ldots\ldots + x_n.$$

$$M_0(t) \text{ of } X = E\{e^{tX}\}$$

(*iv*)

Remark. Since from the 1

generating function.

Ex. 8-31. (*i*) *Discuss the*

(*ii*) *Prove that the r-th cu*

of the r-th cumulants of the v

(*iii*) *Show that* $k_1 = \mu'_1, k$

Sol. (*i*) By Ex. 8-30 (*ii*)

$K_0($

Let $k_1, k_2 \ldots\ldots$ and k'_1, k'_2

(1) $\Rightarrow$ $k_1 t + k_2 \frac{t^2}{2}$

$\therefore$

$\therefore$ Except first all cumu

first) depend on scale.

(*ii*) In Ex. 8-30 (*iii*) takin

$K_0($

$$= E\{e^{t(x_1+...+x_n)}\}$$

$$= E\{e^{tx_1}\, e^{tx_2} ....e^{tx_n}\}$$

$$= E\{e^{tx_1}\}\, E\{e^{tx_2}\}...E\{e^{tx_n}\}$$

($\because$ x_i's are independent)

$$= \{M_0(t) \text{ of } x_1\}...\{M_0(t) \text{ of } x_n\}$$

(*iv*)
$$M_a(t) = E\{e^{t(x-a)}\}$$

$$= E\left\{1 + t(x-a) + \frac{t^2}{2!}(x-a)^2 + ...\right\}$$

$$= 1 + E(x-a) + \frac{t^2}{2!}E(x-a)^2 + ...$$

$$= 1 + t\mu_1'(a) + \frac{t^2}{2!}\mu_2'(a) + ...$$

$$= \sum_{r=0}^{\infty} \mu_r'(a)\frac{t^r}{r!}$$

Remark. Since from the function $M_a(t)$, moments can be generated, it is called moment generating function.

Ex. 8-31. (*i*) *Discuss the effect of change of origin and scale on cumulants.*

(*ii*) *Prove that the r-th cumulant of the sum of independent chance variates is the sum of the r-th cumulants of the variates.*

(*iii*) *Show that* $k_1 = \mu_1', k_2 = \mu_2, k_3 = \mu_3$ and $k_4 = \mu_4 - 3\mu_2^{\,2}$.

Sol. (*i*) By Ex. 8-30 (*ii*)

$$K_0(t) \text{ of } x = \log\{M_0(t) \text{ of } x\}$$

$$= at + \log\{M_0(th) \text{ of } X\}$$

$$= at + K_0(th) \text{ of } X \qquad ...(1)$$

Let k_1, k_2.......and k_1', k_2'...... be the cumulants for x and X respectively. Then

$$(1) \Rightarrow \; k_1 t + k_2\frac{t^2}{2!} + ... = at + k_1'(th) + k_2'\frac{(th)^2}{2!} + ...$$

$$\therefore \qquad k_1 = a + hk_1'$$

$$k_r = h^r k_r' \qquad r \geq 2$$

$\therefore$ Except first all cumulants are independent of origin but all cumulants (including first) depend on scale.

(*ii*) In Ex. 8-30 (*iii*) taking log

$$K_0(t) \text{ of } X = \sum_{i=1}^{n}\{K_0(t) \text{ of } x_i\}$$

Let $k_1, k_2 \ldots\ldots$ and $k_1{}^i, k_2{}^i \ldots\ldots$ be the cumulants of X and x_i respectively $(i = 1, 2 \ldots n)$

$$\therefore \qquad \sum_{r=1} k_r \frac{t^r}{r!} = \sum_{i=1}^{n} \left\{ \sum_{r=1} k_r^i \frac{t^r}{r!} \right\}$$

$$= \sum_{r=1} \frac{t^r}{r!} \left\{ \sum_{i=1}^{n} k_r^i \right\}$$

$$\Rightarrow \qquad k_r = \sum_{i=1}^{n} k_r{}^i$$

(*iii*) By def.,

$$K_{\bar{x}}(t) = \log M_{\bar{x}}(t)$$

$$= \log \left\{ 1 + \left(t\mu_1 + \frac{t^2}{2!}\mu_2 + \ldots\ldots \right) \right\}$$

$$\log \left\{ 1 + \left(\frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3 + \frac{t^4}{4!}\mu_4 \ldots\ldots \right) \right\}$$

$$(\because \mu_1 = 0)$$

$$= \left(\frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3 + \frac{t^4}{4!}\mu_4 + \ldots\ldots \right) - \frac{1}{2}\left(\frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3 + \ldots \right)^2 + \ldots\ldots$$

$$= \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + (\mu_4 - 3\mu_2^2)\frac{t^4}{4!} + \ldots$$

$$\therefore \qquad k_1(\bar{x}) = 0 \Rightarrow k_1(0) - \bar{x} = 0$$

$$\Rightarrow k_1(0) = \bar{x}$$

$$k_2 = \mu_2$$

$$k_3 = \mu_3$$

and $$k_4 = \mu_4 - 3\mu_2{}^2$$

Ex. 8-32. *If x is a variate with zero mean and cumulants k_r, show that the first two cumulants l_1 and l_2 of x^2 are given by*

$$l_1 = k_2, l_2 = 2k_2{}^2 + k_4.$$

Sol. We have for x,

$$k_1 = \mu_1{}'(0) = \bar{x} = 0$$

$$k_2 = \mu_2$$

where μ_r's are moments about r

$\therefore$

Then

$\therefore$ μ_2 of

Ex. 8-33. *Show that*

μ

Sol. $K($

$$\Rightarrow k_1 t + k_2 \frac{t^2}{2!} + \ldots + k_r \frac{t^r}{r!} + \ldots =$$

Differentiating w.r.t. t

$$k_1 + k_2 t + \ldots k_r \frac{t^{r-1}}{(r-1)!} + \ldots = \{$$

$$\Rightarrow \left\{ k_1 + k_2 t + k_3 \frac{t^2}{2!} + \ldots k_r \frac{t}{(r} \right.$$

$$k_3 = \mu_3$$

$$k_4 = \mu_4 - 3\mu_2^{\,2}$$

where μ_r's are moments about mean for x.

$$\therefore \qquad l_1 = \text{first cumulant of } x^2$$

$$= E(x)^2$$

$$= \mu_2 = k_2$$

$$l_2 = \text{second cumulant of } x^2$$

$$= \mu_2 \text{ of } x^2$$

$$let = y = x^2$$

Then $$\bar{y} = E(x^2) = \mu_2 = k_2$$

$$\therefore \qquad \mu_2 \text{ of } x^2 = E(y-\bar{y})^2$$

$$= E(x^2 - k_2)^2$$

$$= E(x^4) - 2k_2 E(x^2) + k_2^{\,2}$$

$$= \mu_4 - 2k_2\bar{y} + k_2^{\,2}$$

$$= \mu_4 - k_2^{\,2}$$

$$= k_4 + 3\mu_2^{\,2} - k_2^{\,2}$$

$$= k_4 + 3k_2^{\,2} - k_2^{\,2}$$

$$= k_4 + 2k_2^{\,2}.$$

Ex. 8-33. *Show that*

$$\mu_r' = \sum_{j=1}^{r} {}^{r-1}C_{j-1,}\, \mu_{r-j}'\, k_j$$

Sol. $$K(t) = \log M_0(t)$$

$$\Rightarrow k_1 t + k_2 \frac{t^2}{2!} + \ldots + k_r \frac{t^r}{r!} + \ldots = \log\left\{1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \ldots + \mu_r' \frac{t^r}{r!} + \ldots\right\}$$

Differentiating w.r.t. t

$$k_1 + k_2 t + \ldots k_r \frac{t^{r-1}}{(r-1)!} + \ldots = \left\{\frac{\mu_1' + \mu_2' t + \ldots + \mu_r' \dfrac{t^{r-1}}{(r-1)} + \ldots}{1 + \mu_1' t + \ldots + \mu_r' \dfrac{t^r}{r!} + \ldots}\right\}$$

$$\Rightarrow \left\{k_1 + k_2 t + k_3 \frac{t^2}{2!} + \ldots k_r \frac{t^{r-1}}{(r-1)!} + \ldots\right\}\left\{1 + \mu_1' t + \ldots + \mu_r' \frac{t^r}{r!} + \ldots\right\}$$

$$= \left\{\mu_1' + \mu_2' t + \ldots + \mu_r' \frac{t^{r-1}}{(r-1)!} + \ldots\right\}$$

Equating Co-efficients of $\frac{t^{r-1}}{(r-1)!}$

$$\mu_r' = k_1\mu_{r-1}' + k_2(r-1)\mu_{r-2}' + k_3\frac{(r-1)(r-2)}{2!}\mu_{r-3}$$

$$+ \ldots\ldots\ldots\ldots + k_r$$

$$= k_1\mu_{r-1}' + {}^{r-1}c_1 k_2\mu_{r-2}' + {}^{r-1}c_2 k_3\mu_{r-3}' + \ldots + {}^{r-1}c_{r-1}k_r$$

$$= \sum_{j=1}^{r} {}^{r-1}C_{j-1} k_j \mu_{r-j}'.$$

8.5. Characteristic Function (c.f.)

The characteristic function of a random variate x is defined to be

$$E\{e^{itx}\}$$

where t is real and $i^2 = -1$. It is denoted by $\phi_x(t)$ or simply $\phi(t)$

$$\therefore \qquad \phi(t) = E(e^{itx})$$

8.5.1. Properties of characteristic Function

c.f. possesses following properties :

(*i*) $\phi(t)$ is defined in every finite interval and is continuous

(*ii*) $\phi(0) = E(e^0) = 1$

(*iii*) $|\phi(t)| = \left|\sum_x e^{itx} p(x)\right|$

$$\leq \sum_x \left|e^{itx}\right| p(x)$$

$$\leq \sum_x p(x) = 1$$

$(\because \left|e^{itx}\right| = |\cos(xt) + i\sin(xt)| = \sqrt{\cos^2(xt) + \sin^2(xt)} = 1)$

(*iv*) $\phi(t)$ and $\phi(-t)$ are conjugate f^ns.

The relation between moments and c.f. is

$$\mu_r' = (-i)^r \left\{\frac{\partial^r}{\partial t^r}\phi(t)\right\}_{t=0}$$

Theorem 8.4.2. If the prob. f^n is symmetrical about 0 *i.e.*, $p(-x) = p(x)$, $\phi(t)$ is real and even f^n of t.

Proof. By def.

$\phi(t)$

since $p(-x)$

$\phi(t)$

Which

Moreover,

$\phi(-t)$

$\Rightarrow \phi$ is an even f^n.

Remark 8.5.3. C.f. has an advar

m.g.f. may or may not exist.

One of the simplest conditions

(*i*) $\phi(t)$ is bounded and conti

(*ii*) $\phi(0) = 1$

(*iii*) $\psi(x.c) = \int_0^c\int_0^c \phi(t-z)e^{ix(t-}$

is real and non-negative for all real.

Remark 8.5.4. If there are two must be identical.

$\therefore$ For a given c.f. $\phi(t)$, there is

is given by

$f(x)$

Result 8.5.5. If x_1, x_2 are inde

$\phi_{x_1+x_2}(t)$

but its converse may not be true.

Result 8.5.6. For Bivariate dist

$\phi_{x_1,x_2}(t_1, t_2)$

Here x_1 and x_2 are independer

$\phi_{x_1,x_2}(t_1, t_2)$

Ex. 8-34. Find the distribution

$\frac{t^{r-1}}{(r-1)!}+\ldots\}$

$\frac{-1)(r-2)}{2!}\mu_{r-3}$

defined to be

nply $\phi(t)$

ntinuous

$t) = 1)$

t 0 *i.e.*, $p(-x) = p(x), \phi(t)$ is real

Proof. By def.

$$\phi(t) = \Sigma e^{itx} p(x) \qquad \ldots(1)$$

since $\quad p(-x) = p(x)$, (1) can be written as

$$\phi(t) = \Sigma\left\{e^{itx} + e^{-itx}\right\} p(x)$$

Which is real.

Moreover,

$$\phi(-t) = \Sigma e^{-itx} p(x)$$

changing x *to* $-y$

$$= \sum_y e^{ity} p(-y)$$

$$= \sum_y e^{ity} p(y) = \phi(t)$$

$\Rightarrow \phi$ is an even f^n.

Remark 8.5.3. C.f. has an advantage over *m.g.f.* in the fact that c.f. always exists whereas *m.g.f.* may or may not exist.

One of the simplest conditions for a given f^n $\phi(t)$ to be a c.f. are :

(*i*) $\phi(t)$ is bounded and continuous

(*ii*) $\phi(0) = 1$

(*iii*) $\psi(x.c) = \int_o^c \int_0^c \phi(t-z) e^{ix(t-z)}\, dtdz$

is real and non-negative for all real x and all $c > 0$.

Remark 8.5.4. If there are two distributions with identical c.f.s. then the distributions must be identical.

$\therefore$ For a given c.f. $\phi(t)$, there is only one distribution. The density f^n of this distribution is given by

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt$$

Result 8.5.5. If x_1, x_2 are independent variables, then

$$\phi_{x_1+x_2}(t) = \phi_{x_1}(t).\phi_{x_2}(t)$$

but its converse may not be true.

Result 8.5.6. For Bivariate distribution (see Chapter 11) c.f. is defined by

$$\phi_{x_1,x_2}(t_1,t_2) = E\{e^{(t_1x_1+t_2x_2)}\}$$

Here x_1 and x_2 are independent if and only if

$$\phi_{x_1,x_2}(t_1,t_2) = \phi_{x_1}(t).\phi_{x_2}(t_2).$$

Ex. 8-34. Find the distribution for which the characteristic f^n is

$$(i)\ e^{-|t|},\ (ii)\ e^{-\frac{1}{2}t^2}$$

Sol. The density f^n is given by

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\,\phi(t)dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} c^{-itx}e^{-|t|}dt$$

$$= \frac{1}{2\pi}\left[\int_{-\infty}^{o} e^{-itx}e^{t}dt + \int_{o}^{\infty} c^{-itx}e^{-t}dt\right]$$

$$= \frac{1}{2\pi}\int_{o}^{\infty} e^{-t}\left(e^{itx}+e^{-itx}\right)dt$$

$$= \frac{1}{\pi}\int_{0}^{\infty} e^{-t}.\cos tx\, dx$$

$$= \frac{1}{\pi}\cdot\frac{1}{1+x^2}$$

(*ii*)

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}.e^{-\frac{1}{2}t^2}dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\{(t+ix)^2+x^2\}}dt$$

Put $t+ix=z$.

$$= \frac{1}{2\pi}e^{-\frac{1}{2}x^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}dz$$

$$= \frac{1}{2\pi}e^{-\frac{1}{2}x^2}.2\int_{0}^{\infty} e^{-\frac{1}{2}z^2}dz$$

$$= \frac{1}{\pi}e^{-\frac{1}{2}x^2}\frac{1}{\sqrt{2}}\int_{o}^{\infty} e^{-y}y^{\frac{1}{2}-1}dy \qquad \text{Put } \frac{z^2}{2}=y$$

$$= \frac{1}{\pi}e^{-\frac{1}{2}x^2}.\frac{1}{\sqrt{2}}.\Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

8.6. Generating Function.

Consider infinite series

$$a_0+a_1t+a_2t^2+\ldots\ldots \tag{1}$$

with real co-efficients and variable t

If in some interval of t, series converges say to $S(t), S(t)$ is called generating function of sequence $\{a_i\}$

e.g., (*i*) the generating function

which is

(2) the generating function of se

which is e^t.

If x be a discrete random variab

series (1) gives

p

This sum function $E\{t^x\}$ is call

$M_0(T)$

is obtained by replaing t be e^T in pr

Ex. 8-35. *Let $P(t)$ be prob. ge*

(*i*) $x-1$ (*ii*) $3x$ (*iii*) $P(x\geq n)$ (*iv*) F

Sol. Let p_i

(*i*) $p\cdot g\cdot f$ *of* $(x-1)$

(*ii*) $p\cdot g\cdot f$ *of* $(3x)$

(*iii*) Let q_n

and $Q(t)$

q_{n+1}

$\therefore$ q_n-q_{n+1}

$\therefore$ $\sum_{n=0}^{\infty}q_nt^n-\sum_{n=0}^{\infty}q_{n+1}t^n$

$\Rightarrow$ $Q(t)-\frac{1}{t}\{Q(t)-q_o\}$

$t)dt$

dt

$dt + \int_0^\infty c^{-itx} e^{-t} dt \Big]$

$+ e^{-itx}\big) dt$

$tx\, dx$

$^{-\frac{1}{2}t^2} dt$

$^{x)^2+x^2)} dt$

Put $t + ix = z.$

$^{-\frac{1}{2}z^2} dz$

$e^{-\frac{1}{2}z^2} dz$

$e^{-y} y^{\frac{1}{2}-1} dy$ Put $\frac{z^2}{2} = y$

$\Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

... (1)

$t), S(t)$ is called generating function

e.g., (*i*) the generating function of sequence (0, 1, 1, 1,...) is sum of the series

$$t + t^2 + t^3 + \ldots.$$

which is $$\frac{t}{1-t}.$$

(2) the generating function of sequence $\left\{\frac{1}{i!}\right\}$ is sum of the series

$$1 + t + \frac{t^2}{2!} + \ldots$$

which is e^t.

If x be a discrete random variable with non-negative integers as values and

$$a_i = P(x = i) = p_i$$

series (1) gives

$$p_0 + p_1 t + p_2 t^2 + \ldots\ldots$$
$$= E\{t^x\}$$

This sum function $E\{t^x\}$ is called **probability generating function.** Obviously

$$M_0(T) = E\{e^{Tx}\}$$

is obtained by replaing t be e^T in prob. generating function.

Ex. 8-35. *Let $P(t)$ be prob. generating function of a random variate x. Find p.g.f. of* (*i*) $x-1$ (*ii*) $3x$ (*iii*) $P(x \geq n)$ (*iv*) $P(x > n+1)$

Sol. Let $$p_i = p(x = i)$$

(*i*) $$p \cdot g \cdot f \text{ of } (x-1) = E\{t^{x-1}\}$$
$$= \frac{1}{t} E(t^x) = \frac{P(t)}{t}$$

(*ii*) $$p \cdot g \cdot f \text{ of } (3x) = E\{t^{3x}\} = E\{t^3)^x\} = P(t^3)$$

(*iii*) Let $$q_n = P(x \geq n)$$
$$= p_n + p_{n+1} + \ldots\ldots..$$

and $$Q(t) = q_0 + q_1 t + q_2 t^2 + \ldots.$$
$$q_{n+1} = p_{n+1} + p_{n+2} + \ldots\ldots.$$

$\therefore$ $$q_n - q_{n+1} = p_n \quad n \geq o$$

$\therefore$ $$\sum_{n=0}^{\infty} q_n t^n - \sum_{n=0}^{\infty} q_{n+1} t^n = \sum_{n=0}^{\infty} p_n t^n$$

$\Rightarrow$ $$Q(t) - \frac{1}{t}\{Q(t) - q_o\} = P(t)$$

$$\Rightarrow \qquad Q(t) = \frac{q_0 - tP(t)}{1-t}$$

Now $q_0 = p_0 + p_1 + \ldots\ldots = 1$

$$\therefore \qquad Q(t) = \frac{1 - t\,P(t)}{1-t}$$

(*iv*) is left as an exercise.

Ex. 8-36. Find $p.g.f.$ of

(*i*) $P(x < n)$ (*ii*) $P(x \le n)$

$$\left\{Ans: \frac{t\,P(t)}{1-t}, \frac{P(t)}{1-t}\right\}$$

Ex. 8-37. *Let P(t) denote the p.g.f. of a random variate x and Q(t) be the p.g.f. of P(x=2n). Show that*

$$Q(t) = \frac{P(t^{\frac{1}{2}}) + P(-t^{\frac{1}{2}})}{2}$$

Sol. Let $P(x = i) = p_i$

Then

$$P(t) = \sum_{n=0}^{\infty} p_n\, t^n$$

and

$$Q(t) = \sum_{n=0}^{\infty} p_{2n}\, t^n$$

$$(\because P(x = 2n) = p_{2n})$$

$$= p_0 + p_2 t + p_4 t^2 + \ldots.$$

$$\therefore \qquad 2Q(t) = \left\{p_0 + p_1 t^{\frac{1}{2}} + p_2 t + p_3 t^{\frac{3}{2}} + p_4 t^2 + \ldots.\right\}$$

$$+\left\{p_0 - p_1 t^{\frac{1}{2}} + p_2 t - p_3 t^{\frac{3}{2}} + p_4 t^2 \ldots..\right\}$$

$$= P(t^{\frac{1}{2}}) + P(-t^{\frac{1}{2}})$$

$$\therefore \qquad Q(t) = \frac{P(t^{\frac{1}{2}}) + P(-t^{\frac{1}{2}})}{2}$$

Ex. 8-38. *Let x denote the number of failures preceding the first success in an indefinite series of independent trials with constant prob. p of success. Find p.g.f. of x*

Sol. By Ex. 8-6

$$P(x = r) = q^r p,$$

$$\therefore \qquad p(t) = E(t^x)$$

which exists only when qt < 1 i.e.

t

1. Let x be a random variate and

$E(x -$

where $V(x)$ stands for variance

2. If 'a' is constant, show that

(*i*) $E(a) = a$

(*ii*) $E(ax) = aE(x)$

(*iii*) $\text{Var}(ax) = a^2\,\text{Var}(x)$

3. Two fair dice are tossed. Find
4. A and B in turn toss an ordinar
If A has first throw, what are th
5. Show that (*i*) $\mu_1 = k_1$, (*ii*) $\sigma =$
6. Four coins are tossed. What is
7. If $P(x = -1) = \frac{1}{4}, P(x = 0) =$

and

$$P(y = -2) = \frac{1}{3}, P(y = 10) = \frac{1}{3}$$

find $E(x + y)$. Further, if x an

8. In an objective type examinati
are four answers out of which

the correct answer and $-\frac{1}{3}$ o

$$\left\{Ans: \frac{t\,P(t)}{1-t}, \frac{P(t)}{1-t}\right\}$$

variate x and Q(t) be the p.g.f. of

$(\because P(x=2n)=p_{2n})$

$+ \ldots.$

$t + p_3 t^{\frac{3}{2}} + p_4 t^2 + \ldots.\Big\}$

$t - p_3 t^{\frac{3}{2}} + p_4 t^2 \ldots..\Big\}$

ing the first success in an indefinite cess. Find p.g.f. of x

$$P(x=r) = q^r p, \qquad r = 0, 1, \ldots\ldots.$$

$$\therefore \quad p(t) = E(t^x)$$

$$= \sum_{r=0}^{\infty} q^r p.t^r$$

$$= p\left\{qt + q^2 t^2 + \ldots\right\}$$

$$= pqt\left\{1 + qt + q^2 t^2 + \ldots\ldots\right\}$$

$$= \frac{pqt}{1-qt}$$

which exists only when qt < 1 i.e.

$$t < \frac{1}{q}.$$

EXERCISES

1. Let x be a random variate and c, a constant. Show that
$$E(x-c)^2 = V(x) + \{E(x)-c\}^2$$
where $V(x)$ stands for variance of x.
2. If 'a' is constant, show that
   (*i*) $E(a) = a$
   (*ii*) $E(ax) = aE(x)$
   (*iii*) $\text{Var}(ax) = a^2\,\text{Var}(x)$
3. Two fair dice are tossed. Find the probability distribution of the total score.
4. A and B in turn toss an ordinary die for a prize of Rs. 44. The first to toss a 'six' wins. If A has first throw, what are their expectations ?
   [Ans. 24; 20]
5. Show that (*i*) $\mu_1 = k_1$, (*ii*) $\sigma = \sqrt{k_2}$, (*iii*) $\gamma_1 = \frac{k_2}{\sigma^3}$, (*iv*) $\gamma_2 = \frac{k_4}{k_2^2}$
6. Four coins are tossed. What is the expectation of the number of heads.
   [Ans. 2]
7. If $P(x=-1) = \frac{1}{4}, P(x=0) = \frac{1}{2}, P(x=1) = \frac{1}{4}$

and

$$P(y=-2) = \frac{1}{3}, P(y=10) = \frac{1}{3}, P(y=4) = \frac{1}{3},$$

find $E(x+y)$. Further, if x and y are independent, find $E(xy)$.

8. In an objective type examination, consisting of 50 questions, for each question there are four answers out of which only one is correct. A candidate scores 1 if he picks up the correct answer and $-\frac{1}{3}$ otherwise. If a candidate makes only a random choice in

respect of each of the 50 questions, find his expected score and the variance of his score.

9. What is the expected number of double birthdays (two or more persons having the same birthday) in a group of n persons ?

10. A coin is tossed until a head appears. Let x denote the number of tosses. Find *p.g.f.* of x.

$$\left[\textbf{Ans. } \frac{t}{2-t}\right]$$

11. Show that

$$\mu_r'(a) = \left\{\frac{d^r}{dt^r} M_a(t)\right\}_{t=0}$$

12. Let x be a random variable having *m.g.f.* $e^{\lambda(e^t-1)}$. Find $E(x)$

Sol.

$$M_0(t) = e^{\lambda(e^t-1)}$$

$$\frac{dM_0(t)}{dt} = e^{\lambda(e^t-1)}.\lambda e^t$$

$$\therefore \quad E(x) = \left\{\frac{dM_o(t)}{dt}\right\}_{t=0} = \lambda$$

Contin

9.1. Continuous Variable

It is the variable which car
following x will be taken as contin

Probability Density Functi

variate value lying in infinitesim
$f(x)$ dx, is called probability den.

The density function $f(x)$ ha.

(i) $f(x) \geq 0 \quad \forall x.$

(ii) $\int f(x) = 1$ *where the*
variable x.

In the following the density

Probability Differential. '
for a variate to lie in the interval

$$P(a \leq x \leq b) =$$

Probability Curve. The cc
curve or simply probability curve

Distribution Function. The

$$F(x) =$$

is called the cumulative distributi
The c.d.f. F(x) has the following

(i) $F'(x) = f(x) \geq 0 \Rightarrow F($
(ii) $F(-\infty) = 0.$
(iii) $F(\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$

It is also sometimes denoted
Any continuous differentiabl
as c.d.f. of x. Then density functi

Indicater Function : Indic

$$I_{(a,\,b)}(x) =$$

Ex. 9-1. If $f_X(\,.\,)$ *denotes*
cumulative distribution function,

ted score and the variance of his

(two or more persons having the

ıote the number of tosses. Find

$\left[\textbf{Ans. } \frac{t}{2-t}\right]$

$t=0$

nd $E(x)$

$= \lambda$

9

Continuous Distributions

9.1. Continuous Variable

It is the variable which can take all possible values between certain limits. In the following x will be taken as continuous variable and will also be used to represent its values.

Probability Density Function. *A continuous function $f(x)$. s.t. the probability of the variate value lying in infinitesimal interval $x - \frac{dx}{2}$ or $x + \frac{dx}{2}$ can be expressed in the form $f(x)\,dx$, is called probability density function or simply the density function.*

The density function $f(x)$ has the following properties :

(*i*) $f(x) \geq 0 \quad \forall x.$

(*ii*) $\int f(x) = 1$ *where the integration is being extended to the entire range of the variable x.*

In the following the density function will be denoted by $f(x)$ {or $f_x(x)$}.

Probability Differential. *'$f(x)\,dx$' is called probability differential. The probability for a variate to lie in the interval (a, b) is given by*

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx.$$

Probability Curve. The continuous curve $y = f(x)$ is called the probability density curve or simply probability curve.

Distribution Function. The function $F(x)$ defined by

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

is called the cumulative distribution function (c.d.f.) or simply the distribution function of x. The c.d.f. F(x) has the following properties :

(*i*) $F'(x) = f(x) \geq 0 \Rightarrow F(x)$ *is non-decreasing function.*

(*ii*) $F(-\infty) = 0.$

(*iii*) $F(\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1.$

It is also sometimes denoted by $F_x(x)$.

Any continuous differentiable function $F(x)$ with the above properties may be regarded as c.d.f. of x. Then density function of x is $F'(x)$.

Indicater Function : Indicater function for continuous variable x is defined by

$$I_{(a,\,b)}(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Ex. 9-1. If $f_X(\,.\,)$ *denotes the p.d.f. of the random variable X and $F_X(\,.\,)$ be its cumulative distribution function, show that*

$$E(X) = \int_0^{\infty} \{1 - F_X(x)\}\,dx - \int_{-\infty}^{0} F_X(x)\,dx.$$

Sol. Consider

$$I_1 = \int_0^{\infty} \{1 - F_X(x)\}\,dx$$

$$= \int_0^{\infty} \{1 - P(X \le x)\}\,dx = \int_0^{\infty} P(X > x)\,dx$$

$$= \int_0^{\infty} \left\{ \int_x^{\infty} f_x(t)\,dt \right\} dx$$

$$= \int_{t=0}^{\infty} f_x(t) \left\{ \int_0^{t} dx \right\} dt \qquad \text{(Changing the order of integration)}$$

$$= \int_0^{\infty} t f_X(t)\,dt$$

$$I_2 = \int_{-\infty}^{0} F_X(x)\,dx = \int_{-\infty}^{0} P(X \le x)\,dx$$

$$= \int_{-\infty}^{0} dx \int_{-\infty}^{x} f_x(t)\,dt = \int_{-\infty}^{0} f_x(t) \left\{ \int_t^{0} dx \right\} dt$$

$$= -\int_{-\infty}^{0} t . f_X(t)\,dt$$

$$\therefore \quad \text{R.H.S.} = I_1 - I_2.$$

$$= \int_0^{\infty} t . f_X(t)\,dt + \int_{-\infty}^{0} t . f_X(t)\,dl$$

$$= \int_{-\infty}^{\infty} t . f_X(t)\,dt = E(X).$$

Ex. 9-2. *For a continuous random variate X, show that*

$$var(X) = \int_0^{\infty} 2x\,\{1 - F_X(x) + F_X(-x)\}\,dx - \overline{X}^2.$$

Sol.

$$I = \int_0^{\infty} 2x\,\{1 - F_X(x) + F_X(-x)\}\,dx$$

$$= \int_0^{\infty} 2x\,\{P(X \ge x) + P(X \le -x)\}\,dx$$

$$= \int_0^{\infty} 2x \left\{ \int_x^{\infty} f_x(y)\,dy \right\} dx + \int_0^{\infty} 2x \left\{ \int_{-\infty}^{-x} f_x(y)\,dy \right\} dx$$

Consider

$$I_1 = \int_0^{\infty} 2x \left\{ \int_x^{\infty} f_X(\right.$$

$$= \int_0^{\infty} f_X(y) \left\{ \int_0^{y} 2 \right.$$

$$= \int_0^{\infty} y^2 f_X(y)\,dy$$

$$= \int_0^{\infty} x^2 f_X(x)\,dx$$

$$I_2 = \int_0^{\infty} 2x \left\{ \int_{-\infty}^{-x} f_X \right.$$

$$= \int_{-\infty}^{0} f_X(y) \left\{ \int_0^{-y} \right.$$

$$= \int_{-\infty}^{0} y^2 f_X(y)\,d$$

$$\therefore \quad I = \int_0^{\infty} x^2 f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,d$$

$$\therefore \quad \text{R.H.S.} = \int_{-\infty}^{\infty} x^2 f_X(x)\,d$$

$$= \text{Var}(X).$$

Mean Moments etc. *All the*

distribution with the difference $\frac{f_i}{N}$

the range of the variate.

Median. Median '**a**' is give

$$\int_{-\infty}^{a} f(x)\,dx = \int_a^{\infty}$$

Quartiles. *The lower and u*

$$\int_{-\infty}^{Q_1} f(x)\,dx = \frac{1}{4}$$

Mode. *Mode is that value o*

$$f'(x) = 0$$
$$f''(x) < 0$$

provided that the solution of $f'(x$

Ex. 9-3. *Show that for the r*

$$dF = dx$$

$)\,dx.$

$(X > x)\,dx$

[h]anging the order of integration)

x

$\left\{\int_t^0 dx\right\} dt$

t

$x - \overline{X}^2.$

dx

$\}\,dx$

$\left\{\int_{-\infty}^{-x} f_X(y)\,dy\right\} dx$

Consider

$$I_1 = \int_0^\infty 2x \left\{ \int_x^\infty f_X(y)\,dy \right\} dx$$

$$= \int_0^\infty f_X(y) \left\{ \int_0^y 2x\,dx \right\} dy$$

$$= \int_0^\infty y^2 f_X(y)\,dy$$

$$= \int_0^\infty x^2 f_X(x)\,dx$$

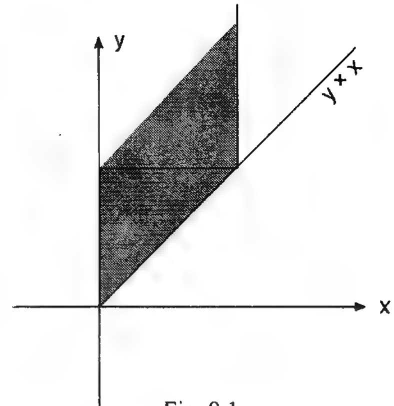


Fig. 9.1.

$$I_2 = \int_0^\infty 2x \left\{ \int_{-\infty}^{-x} f_X(y)\,dy \right\} dx$$

$$= \int_{-\infty}^0 f_X(y) \left\{ \int_0^{-y} 2x\,dx \right\} dy$$

$$= \int_{-\infty}^0 y^2 f_X(y)\,dy = \int_{-\infty}^0 x^2 f_X(x)\,dx$$

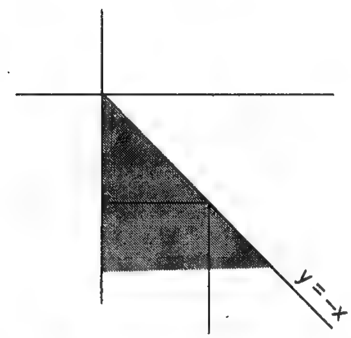


Fig. 9.2.

$$\therefore \quad I = \int_0^\infty x^2 f_X(x)\,dx + \int_{-\infty}^0 x^2 f_X(x)\,dx$$

$$= \int_{-\infty}^\infty x^2 f_X(x)\,dx$$

$$\therefore \quad \text{R.H.S.} = \int_{-\infty}^\infty x^2 f_X(x)\,dx - \overline{X}^2$$

$$= \text{Var}(X).$$

Mean Moments etc. *All the quantities e.g., mean, moments are defined as for discrete distribution with the difference* $\frac{f_i}{N}$ *is replaced by* $f(x)\,dx$ *and summation by integration over the range of the variate.*

Median. Median '**a**' is given by

$$\int_{-\infty}^a f(x)\,dx = \int_a^\infty f(x)\,dx$$

Quartiles. *The lower and upper quartiles* Q_1 and Q_3 are given by

$$\int_{-\infty}^{Q_1} f(x)\,dx = \frac{1}{4} = \int_{Q_3}^\infty f(x)\,dx$$

Mode. *Mode is that value of x for which f(x) is maximum i.e., model value x is s.t.*

$$f'(x) = 0$$
$$f''(x) < 0$$

provided that the solution of $f'(x) = 0$ *lies within the permissible range of* x.

Ex. 9-3. *Show that for the rectangular population*

$$dF = dx \qquad 0 \le x \le 1$$

$$\mu'_1(0) = \frac{1}{2} \text{ and } \mu_2 = \frac{1}{12}.$$

Sol. By def. $\mu'_1(0) = \int_0^1 x\,dx = \frac{1}{2}$

$$\mu'_2(0) = \int_0^1 x^2\,dx = \frac{1}{3}$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \{\mu_1'(0)\} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Ex. 9-4. *For the rectangular distribution* $y = \frac{1}{2a}, -a \le x \le a$. *Show that*

$$M_0(t) = \frac{1}{at} \sin h\, at \text{ and } \mu_{2n} = \frac{a^{2n}}{2n+1}.$$

Sol. By def.

$$M_0(t) = \int_{-a}^{a} e^{tx} \frac{1}{2a}\,dx = \frac{1}{2at}\{e^{tx}\}_{-a}^{a}$$

$$= \frac{1}{2at}\{e^{at} - e^{-at}\} = \frac{\sin h\, at}{at}$$

Also $\mu_{2n} = \int_{-a}^{a} (x-\bar{x})^{2n} \frac{1}{2a}\,dx$

where $\bar{x} = A.M. = \int_{-a}^{a} x \frac{1}{2a}\,dx = 0$

$$\therefore \quad \mu_{2n} = \frac{1}{2a}\int_{-a}^{a} x^{2n}\,dx = \frac{1}{a}\int_0^a x^{2n}\,dx = \frac{a^{2n}}{2n+1}.$$

Ex. 9-5. *Calculate* β_1 *for the dist.* $dF = kxe^{-x}\,dx, 0 < x < \infty$.

Sol. k is given by

$$k\int_0^\infty xe^{-x}dx = 1$$

or $k\left\{\left|-e^{-x}.x\right|_0^\infty + \int_0^\infty e^{-x}\,dx\right\} = 1$

or $k\left\{-e^{-x}\right\}_0^\infty = 1$

$$\therefore \quad k = 1$$

$$\bar{x} = \int_0^\infty x^2 e^{-x}\,dx = \left\{\left|-x^2 e^{-x}\right|_0^\infty + 2\int_0^\infty xe^{-x}\,dx\right\}$$

$$= 2\left\{\left|-xe^{-x}\right|_0^\infty + \int_0^\infty e^{-x}\,dx\right\} = 2$$

$\mu'_2(0) = \int_0^\infty$

$\mu'_3(0) = \int_0^\infty$

$\therefore \quad \mu_2 = \mu'$

$\mu_3 = \mu'$

$= 2$

$\beta_1 = \mu$

Ex. 9-6. *Find the s.d., har*

$f(x) = 6$

Sol. $\bar{x} = 6$

$\mu'_2(0) = 6$

$\therefore \quad \mu_2 = \frac{}{1}$

$\therefore \quad s.d. = \sqrt{}$

H.M. is given by

$\frac{1}{H} = 6$

$= 6$

$\therefore \quad H = \frac{1}{3}$

To find mode put $f'(x) = 0$

i.e., $1 - 2x = 0$

or $x = \frac{1}{2}$

Since $f''(x) = -12 < 0, x =$

Let a be the median.

Then $6\int_0^a (x - x^2)dx =$

or $\frac{a^2}{2} - \frac{a^3}{3} = \frac{}{1}$

or $4a^3 - 6a^2 + 1 = 0$

$\therefore \quad a =$

$$\mu'_2(0) = \int_0^\infty x^3 e^{-x}\,dx = \left|-e^{-x}x^3\right|_0^\infty + 3\int_0^\infty x^2 e^{-x}\,dx = 6$$

$$\mu'_3(0) = \int_0^\infty x^4 e^{-x}\,dx = 4\int_0^\infty x^3 e^{-x}\,dx = 24$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \bar{x}^2 = 6 - 4 = 2$$

$$\mu_3 = \mu'_3(0) - 3\mu'_2(0)\,\bar{x} + 2\bar{x}^3$$

$$= 24 - 36 + 16 = 4$$

$$\beta_1 = \mu_3^2/\mu_2^3 = 2.$$

Ex. 9-6. *Find the s.d., harmonic mean, the mode and the median of the dist. given by*

$$f(x) = 6(x - x^2),\ 0 \le x \le 1.$$

Sol.

$$\bar{x} = 6\int_0^1 x(x - x^2)\,dx = 6\left\{\frac{1}{3} - \frac{1}{4}\right\} = \frac{1}{2}$$

$$\mu'_2(0) = 6\int_0^1 x^2(x - x^2)\,dx = 6\left\{\frac{1}{4} - \frac{1}{5}\right\} = \frac{3}{10}$$

$$\therefore \quad \mu_2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}$$

$$\therefore \quad s.d. = \sqrt{\frac{1}{20}}$$

H.M. is given by

$$\frac{1}{H} = 6\int_0^1 \frac{1}{x}(x - x^2)\,dx = 6\int_0^1 (1 - x)\,dx$$

$$= 6\left(1 - \frac{1}{2}\right) = 3$$

$$\therefore \quad H = \frac{1}{3}$$

To find mode put $f'(x) = 0$

i.e., $\quad 1 - 2x = 0$

or $\quad x = \frac{1}{2}$

Since $f''(x) = -12 < 0$, $x = \frac{1}{2}$ is the mode

Let a be the median.

$$\text{Then } 6\int_0^a (x - x^2)\,dx = \frac{1}{2}$$

or $\quad \frac{a^2}{2} - \frac{a^3}{3} = \frac{1}{12}$

or $\quad 4a^3 - 6a^2 + 1 = 0$

$$\therefore \quad a = \frac{1}{2}.$$

Ex. 9-7. *For the dist.*

$$dF = y_0 e^{-|x|}\,dx, \; -\infty < x < \infty$$

show that $y_0 = \frac{1}{2}$, $\mu'_1(0) = 0$, $\sigma = \sqrt{2}$ *and mean deviation about mean* $= 1$.

Sol. y_0 is given by

$$y_0 \int_{-\infty}^{\infty} e^{-|x|}\,dx = 1$$

or $$2y_0 \int_{0}^{\infty} e^{-|x|}\,dx = 1 \qquad (\because e^{-|x|} \text{ is an even } f^n \text{ of } x)$$

$$\therefore \quad 2y_0 \int_{0}^{\infty} e^{-x}\,dx = 1 \qquad (\because |x| = x \text{ as } x \geq 0)$$

$$\therefore \quad y_0 = \frac{1}{2}$$

$$\mu'_1(0) = \frac{1}{2}\int_{-\infty}^{\infty} x e^{-|x|}\,dx = 0$$

$$\therefore \quad \mu_2 = \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|}\,dx = \int_{0}^{\infty} x^2 e^{-x}\,dx$$

$$= \left|-x^2 e^{-x}\right|_0^{\infty} + 2\int_0^{\infty} x e^{-x}\,dx = 2\int_0^{\infty} x e^{-x}\,dx$$

$$= 2\left\{\left|-xe^{-x}\right|_0^{\infty} + \int_0^{\infty} e^{-x}\,dx\right\} = 2$$

$$\therefore \quad \sigma = \sqrt{2}$$

$$\text{Mean deviation about mean} = \frac{1}{2}\int_{-\infty}^{\infty} |x - 0|\, e^{-|x|}\,dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} |x| e^{-|x|}\,dx = \int_0^{\infty} x e^{-x}\,dx = 1.$$

Ex. 9-8. *Show that for the dist.*

$$f(x) = \frac{2a}{\pi}\left(\frac{1}{a^2 + x^2}\right), \; -a \leq x \leq a$$

$$\mu_2 = a^2\,\frac{(4-\pi)}{\pi}, \; \mu_4 = a^4\left(1 - \frac{8}{3\pi}\right).$$

Sol. $$\int_{-a}^{a} f(x)\,dx = \frac{2a}{\pi}\int_{-a}^{a} \frac{1}{a^2 + x^2}\,dx = \frac{2a}{\pi}\left|\frac{1}{a}\tan^{-1}\frac{x}{a}\right|_{-a}^{a} = 1$$

$$\mu'_1(0) = \frac{2a}{\pi}\int_{-a}^{a} x\,\frac{1}{a^2 + x^2}\,dx = 0$$

$$\mu_2 = \frac{2a}{\pi}\int_{-a}^{a} x^2\,\frac{1}{a^2 + x^2}\,dx = \frac{4a}{\pi}\int_0^{a} \frac{x^2 + a^2 - a^2}{a^2 + x^2}\,dx$$

= 4 …

= a …

μ_4 = 2 …

= 4 …

Ex. 9-9. *For a continuous*

$f(x)$ = 3 …

Find the first three moment…

symmetrical about the mean wi…

Sol. $\mu'_1(0) = \frac{3}{4}$ …

$\mu'_2(0) = \frac{3}{4}$ …

$\mu'_3(0) = \frac{3}{4}$ …

$\therefore$ $\mu_2 = \mu$ …

Let 'a' be the median

Then $\int_0^a f(x)\,dx = \frac{1}{2}$

or $\frac{3}{4}\int_0^a x(2-x)dx = \frac{1}{2}$

or $a^2 - \frac{a^3}{3} = \frac{2}{3}$

or $a^3 - 3a^2 + 2 = 0$

or $(a-1)(a^2 - 2a - 2)$ …

$\therefore$ $a = 1$

$\therefore$ Median = Mean

$\therefore$ Dist. is symmetrical …

Ex. 9-10. *Find the mean d…*

$f(x) = \frac{3}{4}$ …

Sol. From last example, D…

$\therefore$ Skewness = 0

about mean = 1.

($\because e^{-|x|}$ is an even f^n of x)

($\because |x| = x$ as $x \geq 0$)

$\int_0^\infty xe^{-x}\,dx$

2

1.

$^{-1}\frac{x}{a}\Big|_{-a}^{a} = 1$

$\frac{x^2+a^2-a^2}{a^2+x^2}\,dx$

$$= \frac{4a}{\pi}\int_0^a \left\{1-\frac{a^2}{a^2+x^2}\right\}dx = \frac{4a}{\pi}\left\{a-a\cdot\frac{\pi}{4}\right\}$$

$$= \frac{a^2}{\pi}(4-\pi)$$

$$\mu_4 = \frac{2a}{\pi}\int_{-a}^{a}\left\{x^4\frac{1}{a^2+x^2}\right\}dx = \frac{4a}{\pi}\int_0^a\left\{x^2-a^2+\frac{a^4}{x^2+a^2}\right\}dx$$

$$= \frac{4a}{\pi}\left\{\frac{a^3}{3}-a^3+a^3\frac{\pi}{4}\right\} = a^4\left\{1-\frac{8}{3\pi}\right\}.$$

Ex. 9-9. *For a continuous distribution whose relative frequency density is given by*

$$f(x) = \frac{3x(2-x)}{4};\ 0 \leq x \leq 2.$$

Find the first three moments about the origin. Hence or otherwise show that the dist. is symmetrical about the mean with variance $= \frac{1}{5}$.

Sol.
$$\mu'_1(0) = \frac{3}{4}\int_0^2 x^2(2-x)dx = \frac{3}{4}\left\{\frac{2}{3}x^3-\frac{x^4}{4}\right\}_0^2 = 1$$

$$\mu'_2(0) = \frac{3}{4}\int_0^2 x^3(2-x)\,dx = \frac{3}{4}\left\{\frac{1}{2}x^4-\frac{x^5}{5}\right\}_0^2 = \frac{6}{5}$$

$$\mu'_3(0) = \frac{3}{4}\int_0^2 x^4(2-x)\,dx = \frac{3}{4}\left\{\frac{2}{5}x^5-\frac{x^6}{6}\right\}_0^2 = \frac{8}{5}$$

$$\therefore\quad \mu_2 = \mu'_2(0)-\{\mu'_1(0)\}^2 = \frac{6}{5}-1 = \frac{1}{5}$$

Let 'a' be the median

Then $\int_0^a f(x)\,dx = \frac{1}{2}$

or $\frac{3}{4}\int_0^a x(2-x)dx = \frac{1}{2}$

or $a^2-\frac{a^3}{3} = \frac{2}{3}$

or $a^3-3a^2+2 = 0$

or $(a-1)(a^2-2a-2) = 0$

$\therefore$ $a = 1$

$\therefore$ Median = Mean

$\therefore$ Dist. is symmetrical about mean.

Ex. 9-10. *Find the mean deviation, s.d. and skewness of the dist. given by*

$$f(x) = \frac{3}{4}x(2-x),\ 0 \leq x \leq 2.$$

Sol. From last example, Dist. is symmetrical about mean

$\therefore$ Skewness = 0

$$\text{Mean deviation about mean} = \frac{3}{4}\int_0^2 |x-1|x\,(2-x)\,dx$$

$$= \frac{3}{4}\int_0^1 x(1-x)\,(2-x)\,dx + \frac{3}{4}\int_1^2 x\,(x-1)\,(2-x)\,dx$$

$$= \frac{3}{4}\int_0^1 \{x^3-3x^2+2x\}\,dx + \frac{3}{4}\int_1^2 \{-x^3+3x^2-2x\}\,dx$$

$$= \frac{3}{4}\left\{\frac{1}{4}-1+1\right\}+\frac{3}{4}\left\{\frac{1}{4}\right\} = \frac{3}{8}$$

For *s.d.* see last example.

Ex. 9-11. *In Ex. 9-9 calculate* $\mu'_3(0)$ *and* $\mu'_4(0)$ *and deduce* β_2.

Sol.
$$\mu'_4(0) = \frac{3}{4}\int_0^2 x^5\,(2-x)\,dx = \frac{3}{4}\left\{\frac{x^6}{3}-\frac{x^7}{7}\right\}_0^2 = \frac{16}{7}$$

$$\therefore \quad \mu_4 = \mu'_4(0) - 4\mu'_3(0)\,\mu'_1(0) + 6\mu'_2(0)\,\{\mu'_1(0)\}^2 - 3\{\mu'_1(0)\}^4$$

$$= \frac{16}{7}-4\frac{8}{5}+6\frac{6}{5}-3=\frac{3}{35}$$

$$\therefore \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\frac{3}{35}}{\frac{1}{25}} = \frac{15}{7}.$$

Ex. 9-12. *For the continuous distribution*

$$dF = y_0\,x(2-x)dx, \quad 0 \le x \le 2$$

show that
$$\mu_{2n+1} = 0$$
for each natural number n.

Sol. y_0 is given by

$$1 = y_0\int_0^2 x(2-x)\,dx$$

$$= y_0\left|x^2-\frac{x^3}{3}\right|_0^2$$

$$= y_0\left\{\frac{4}{3}\right\}$$

$$\Rightarrow \quad y_0 = \frac{3}{4}$$

$$\therefore \quad dF = \frac{3}{4}\,x(2-x)\,dx, \qquad 0 \le x\ 2$$

As in Ex. 9-9 $\quad \bar{x} = 1.$

$$\therefore \quad \mu_{2n+1} = E\{x-1\}^{2n+1}$$

$$= \frac{3}{4}\int_0^2 (x-1)^{2n+1}\,x(2-x)\,dx$$

Put $\quad x-1 = y.$

$= \frac{3}{4}\int^1$

$= \frac{3}{4}\int^1$

$= 0$

Ex. 9-13. *For the distributio*

$f(x) = 1-$

find mean, s.d, β_1 and β_2.

Sol. $\int_0^2 f(x)\,dx = \int_0^2 \{$

Put $1-x = y$

$= \int_1^{-1}$

$= 2\int_0^1$

$= 2\{$

Mean $= E(x$

$= \int_0^2$

Put $1-x = y$

$= \int_1^{-1}$

$= 2\int_0^1$

$= 2\{$

$\mu'_2(0) = \int_0^2$

$= \int_1^{-1}$

$= \int_{-1}^1$

$= 2\int^1$

$) \, dx$

$x) \, dx + \frac{3}{4} \int_1^2 x(x-1)(2-x) \, dx$

$x\} \, dx + \frac{3}{4} \int_1^2 \{-x^3 + 3x^2 - 2x\} \, dx$

$\left\{\frac{1}{4}\right\} = \frac{3}{8}$

nd deduce β_2.

$\left. \frac{6}{} - \frac{x^7}{7} \right\}_0^2 = \frac{16}{7}$

$\mu'_2(0) \, \{\mu'_1(0)\}^2 - 3\{\mu'_1(0)\}^4$

$0 \le x \; 2$

dx

$$= \frac{3}{4} \int_{-1}^{1} y^{2n+1} (y+1)(1-y) \, dy$$

$$- \frac{3}{4} \int_{-1}^{1} y^{2n+1} (1-y^2) \, dy$$

$$= 0 \qquad (\because \; f^n \text{ to be integrated is odd}).$$

Ex. 9-13. *For the distribution*

$$f(x) = 1 - |1-x|, \; 0 \le x \le 2$$

find mean, s.d, β_1 and β_2.

Sol. $$\int_0^2 f(x) \, dx = \int_0^2 \{1 - |1-x|\} \, dx$$

Put $$1 - x = y$$

$$= \int_1^{-1} \{1 - |y|\} (-dy)$$

$$= 2\int_0^1 (1-y) \, dy$$

$$= 2\left\{y - \frac{y^2}{2}\right\}_0^1 = 2\left(1 - \frac{1}{2}\right) = 1$$

$$\text{Mean} = E(x)$$

$$= \int_0^2 x\{1 - |1-x|\} \, dx$$

Put $$1 - x = y$$

$$= \int_1^{-1} (1-y)\{1 - |y|\}(-dy)$$

$$= 2\int_0^1 \{1-y\} \, dy - \int_{-1}^{1} y\{1-|y|\} \, dy$$

$$= 2\left\{1 - \frac{1}{2}\right\} = 1$$

$$\mu'_2(0) = \int_0^2 x^2 \{1 - |1-x|\} \, dx$$

$$= \int_1^{-1} (1-y)^2 \{1 - |y|\} (-dy), \qquad y = 1 - x$$

$$= \int_{-1}^{1} (1 - 2y + y^2)\{1 - |y|\} dy$$

$$= 2\int_0^1 (1+y^2)(1-y) \, dy$$

$$= 2\int_0^1 \{1 - y + y^2 - y^3\}\, dy$$

$$= 2\left\{1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4}\right\} = \frac{7}{6}$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \bar{x}^2$$

$$= \frac{7}{6} - 1 = \frac{1}{6}$$

$$\therefore \quad \text{s.d.} = \frac{1}{\sqrt{6}}.$$

$$\mu'_3(0) = \int_0^2 x^3 \{1 - |1 - x|\}\, dx$$

$$= \int_1^{-1} (1 - y)^3 \{1 - |y|\} (-dy)$$

$$= \int_{-1}^1 (1 - 3y + 3y^2 - y^3) \{1 - |y|\}\, dy$$

$$= 2\int_0^1 (1 + 3y^2)(1 - y)\, dy$$

$$= 2\int_0^1 \{1 - y + 3y^2 - 3y^3\}\, dy$$

$$= 2\left\{1 - \frac{1}{2} + 3\cdot\frac{1}{3} - 3\cdot\frac{1}{4}\right\} = \frac{3}{2}$$

$$\therefore \quad \mu_3 = \mu'_3(0) - 3\mu'_2(0)\,\mu'_1(0) + 2\{\mu'_1(0)\}^3$$

$$= \frac{3}{2} - 3\cdot\frac{7}{6}\cdot 1 + 2 = 0$$

$$\therefore \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$$

$$\mu'_4(0) = \int_0^1 x^4 \{1 - |1 - x|\}\, dx$$

$$= \int_1^1 (1 - y)^4 \{1 - |y|\} (-dy)$$

$$= \int_{-1}^1 (1 - 4y + 6y^2 - 4y^3 + y^4) \{1 - |y|\}\, dy$$

$$= 2\int_0^1 (1 + 6y^2 + y^4)\{1 - y\}\, dy$$

$$= \frac{31}{15}$$

$$\mu_4 = \mu'_4(0)$$

$$= \frac{31}{15} - \ldots$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \ldots$$

Ex. 9-14. *Show that for the dist... to σ and the inter-quartile range is ...*

Sol. The constant y_0 is given by

$$y_0 \int_0^\infty e^{-x/\sigma}\, dx = 1 \text{ or } \ldots$$

$$\therefore \quad \bar{x} = y_0 \int_0^\infty x \ldots$$

$$= \sigma$$

$$\mu'_2(0) = y_0 \int_0^\infty \ldots$$

$$= 2\sigma^2$$

$$\mu_2 = 2\sigma^2 - \ldots$$

Let Q_1 and Q_3 be the quartile ...

Then $y_0 \int_0^{Q_1} e^{-x/\sigma}\, dx = \frac{1}{4}$ and ...

or $\quad \frac{1}{\sigma}\left\{-\sigma e^{-x/\sigma}\right\}_0^{Q_1} = \frac{1}{4}$ an...

or $\quad 1 - e^{-Q_1/\sigma} = \frac{1}{4}$ an...

or $\quad Q_1 = \sigma \log_e \ldots$

$$\therefore \quad Q_3 - Q_1 = \sigma \log_e \ldots$$

Ex. 9-15. *Prove that the geome...*

$$dF = 6(2 - \ldots$$

is given by $6 \log (16G) = 19$.

Sol.

$$\log G = 6\int_1^2 \ldots$$

$$= 6\int_0^1 y \ldots$$

$$= 6\left[\left|\left(\ldots\right.\right.\right.$$

$$\mu_4 = \mu'_4(0) - 4\mu'_3(0)\,\mu'_1(0) + 6\mu'_2(0)\,\{\mu'_1(0)\}^2 - 3\,\{\mu'_1(0)\}^4$$

$$= \frac{31}{15} - 4\cdot\frac{3}{2}\cdot 1 + 6\cdot\frac{7}{6}\cdot 1 - 3 = \frac{1}{15}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{1/15}{1/36} = \frac{12}{5}.$$

Ex. 9-14. *Show that for the dist.* $dF = y_0\, e^{-x/\sigma}\, dx,\ 0 < x < \infty$ *the mean and s.d are equal to* σ *and the inter-quartile range is* $\sigma \log_e 3$.

Sol. The constant y_0 is given by

$$y_0 \int_0^\infty e^{-x/\sigma}\, dx = 1 \quad\text{or}\quad y_0 = \frac{1}{\sigma}$$

$$\therefore\quad \bar{x} = y_0 \int_0^\infty x e^{-x/\sigma}\, dx = \frac{1}{\sigma}\int_0^\infty x e^{-x/\sigma}\, dx = \sigma \int_0^\infty y e^{-y}\, dy$$

where $y = \dfrac{x}{\sigma}$

$$= \sigma$$

$$\mu'_2(0) = y_0 \int_0^\infty x^2 e^{-x/\sigma}\, dx = \sigma^2 \int_0^\infty y^2 e^{-y}\, dy = \sigma^2\, 2!$$

$$= 2\sigma^2$$

$$\mu_2 = 2\sigma^2 - \sigma^2 = \sigma^2 \qquad \therefore \text{ s.d.} = \sigma$$

Let Q_1 and Q_3 be the quartile

$$\text{Then } y_0 \int_0^{Q_1} e^{-x/\sigma}\, dx = \frac{1}{4} \quad\text{and}\quad y_0 \int_0^{Q_3} e^{-x/\sigma}\, dx = \frac{3}{4}$$

or
$$\frac{1}{\sigma}\left\{-\sigma e^{-x/\sigma}\right\}_0^{Q_1} = \frac{1}{4} \quad\text{and}\quad \frac{1}{\sigma}\left\{-\sigma e^{-x/\sigma}\right\}_0^{Q_3} = \frac{3}{4}$$

or
$$1 - e^{-Q_1/\sigma} = \frac{1}{4} \quad\text{and}\quad 1 - e^{-Q_3/\sigma} = \frac{3}{4}$$

or
$$Q_1 = \sigma \log_e \frac{4}{3} \quad\text{and}\quad Q_3 = \sigma \log_e 4$$

$$\therefore\quad Q_3 - Q_1 = \sigma \log_e 3.$$

Ex. 9-15. *Prove that the geometric mean G of the dist.*

$$dF = 6(2-x)(x-1)\, dx,\ 1 \le x \le 2$$

is given by $6 \log (16G) = 19$.

Sol.
$$\log G = 6 \int_1^2 \log x.\,(2-x)(x-1)dx$$

$$= 6 \int_0^1 y(-y+1)\log(1+y)\, dy \qquad \text{where } x - 1 = y$$

$$= 6\left[\left|\left(\frac{y^2}{2} - \frac{y^3}{3}\right)\log(y+1)\right|_0^1 - \int_0^1 \frac{1}{y+1}\left\{\frac{y^2}{2} - \frac{1}{3}y^3\right\}dy\right]$$

$$= \log 2 - 3\int_0^1 \left\{y - 1 + \frac{1}{y+1}\right\}dy + 2\int_0^1 \left(y^2 - y + 1 - \frac{1}{y+1}\right)dy$$

$$= \log 2 - 3\left\{\frac{y^2}{2} - y\right\}_0^1 + 2\left\{\frac{y^3}{3} - \frac{y^2}{2} + y\right\}_0^1 - 5\left|\log(y+1)\right|_0^1$$

$$= \frac{19}{6} - 4\log 2$$

$\therefore \quad 6\log(16\,G) = 19.$

Ex. 9-16. *The elementary probability law of a continuous random variable x is $p(x) = y_0\, e^{-b(x-a)}$ $a \le x < \infty$, where a, b, y_0 are constants. Show that $y_0 = b = \dfrac{1}{\sigma}$ and $a = m - \sigma$ where m, σ are respectively the mean and the s.d. of the dist. Show also that $\beta_1 = 4$ and $\beta_2 = 9$.*

Sol. y_0 is given by

$$y_0 \int_a^\infty e^{-b(x-a)}dx = 1$$

or
$$y_0\left\{\frac{e^{-b(x-a)}}{-b}\right\}_a^\infty = 1 \qquad \therefore\ y_0 = b$$

$$m = y_0\int_a^\infty x e^{-b(x-a)}dx = y_0\left\{\left|\frac{xe^{-b(x-a)}}{-b}\right|_a^\infty + \frac{1}{b}\int_a^\infty e^{-b(x-a)}dx\right\}$$

$$= y_0\,\frac{a}{b} + \frac{y_0}{b^2}\left\{-e^{-b(x-a)}\right\}_a^\infty = a + \frac{1}{b}$$

$$\mu'_2(0) = y_0\int_a^\infty x^2 e^{-b(x-a)}dx$$

$$= y_0\left\{\left|-\frac{x^2}{b}e^{-b(x-a)}\right|_a^\infty + \frac{2}{b}\int_a^\infty xe^{-b(x-a)}dx\right\}$$

$$= y_0\,\frac{a^2}{b} + \frac{2}{b}\,y_0\int_a^\infty xe^{-b(x-a)}dx$$

$$= a^2 + \frac{2}{b}\left(a + \frac{1}{b}\right)$$

$$\therefore \quad \mu_2 = a^2 + \frac{2}{b}\left(a + \frac{1}{b}\right) - \left(a + \frac{1}{b}\right)^2 = \frac{1}{b^2}$$

$$\therefore \quad \sigma^2 = \frac{1}{b^2}$$

$$\therefore \quad y_0 = b = \frac{1}{\sigma}$$

and
$$m - \sigma = a$$

$$\mu'_3(0) = y_0\int_a^\infty x^3 e^{-b(x-a)}dx$$

$$\cdot dy + 2\int_0^1 \left(y^2 - y + 1 - \frac{1}{y+1}\right) dy$$

$$\frac{y^3}{3} - \frac{y^2}{2} + y\Bigg\}_0^1 - 5\left|\log(y+1)\right|_0^1$$

:ontinuous random variable x is

. *Show that* $y_0 = b = \frac{1}{\sigma}$ *and* $a = m$

the dist. Show also that $\beta_1 = 4$ *and*

$$\frac{xe^{-b(x-a)}}{-b}\Bigg|_a^\infty + \frac{1}{b}\int_a^\infty e^{-b(x-a)}\,dx\Bigg\}$$

$$= a + \frac{1}{b}$$

$$\frac{2}{b}\int_a^\infty xe^{-b(x-a)}\,dx\Bigg\}$$

$^{a)}\ dx$

$$= \frac{1}{b^2}$$

$$= y_0\left\{\left|-\frac{x^3}{b}e^{-b(x-a)}\right|_a^\infty + \frac{3}{b}\int_a^\infty x^2 e^{-b(x-a)}\,dx\right\}$$

$$= a^3 + \frac{3}{b}\left\{a^2 + \frac{2}{b}\left(a + \frac{1}{b}\right)\right\}$$

$$\therefore \quad \mu_3 = \mu'_3(0) - 3\mu'_2(0)\,\mu'_1(0) + 2\{\mu'_1(0)\}^3$$

$$= a^3 + \frac{3a^2}{b} + \frac{6}{b^2}\left(a + \frac{1}{b}\right) - 3\left\{a^2 + \frac{2}{b}\left(a + \frac{1}{b}\right)\right\}\left(a + \frac{1}{b}\right)$$

$$+ 2\left(a + \frac{1}{b}\right)^3 = \frac{2}{b^3}$$

$$\therefore \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 4$$

$$\mu'_4(0) = y_0\int_a^\infty x^4 e^{-b(x-a)}\,dx$$

$$= y_0\left\{\left|-\frac{x^4}{b}e^{-b(x-a)} + \frac{4}{b}\int_a^\infty x^3 e^{-b(x-a)}\,dx\right\}\right.$$

$$= a^4 + \frac{4}{b}\left\{a^3 + 3\frac{a^2}{b} + \frac{6}{b^2}\left(a + \frac{1}{b}\right)\right\}$$

$$\therefore \quad \mu_4 = \mu'_4(0) - 4\mu'_3(0)\,\mu'_1(0) + 6\mu'_2(0)\,\{\mu'_1(0)\}^2 - 3\{\mu'_1(0)\}^4$$

$$= a^4 + \frac{4}{b}\left\{a^3 + \frac{3a^2}{b} + \frac{6a}{b^2} + \frac{6}{b^3}\right\}$$

$$-4\left\{a^3 + \frac{3a^2}{b} + \frac{6a}{b^2} + \frac{6}{b^3}\right\}\left(a + \frac{1}{b}\right) +$$

$$6\left\{a^2 + \frac{2a}{b} + \frac{2}{b^2}\right\}\left(a + \frac{1}{b}\right)^2 - 3\left(a + \frac{1}{b}\right)^4$$

$$= \frac{9}{b^4}$$

$$\therefore \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = 9.$$

Ex. 9-17. *For the dist.*

$$dF = \frac{1}{\beta(m,n)}\,x^{n-1}(1-x)^{m-1}\,dx \quad 0 \le x \le 1$$

find the mean, s.d. and the harmonic mean.

Sol. $$\bar{x} = \frac{1}{\beta(m,n)}\int_0^1 x^{n+1-1}(1-x)^{m-1}\,dx = \frac{\beta(n+1,m)}{\beta(m,n)}$$

$$= \frac{|\overline{n+1}\,|\overline{m}}{|\overline{n+m+1}} \cdot \frac{|\overline{n+m}}{|\overline{n}\,|\overline{m}} = \frac{n}{n+m}$$

$$\mu'_2(0) = \frac{1}{\beta(m,n)} \int_0^1 x^{n+2-1}(1-x)^{m-1}\,dx = \frac{\beta(n+2,m)}{\beta(m,n)}$$

$$= \frac{|\overline{n+2}\,|\overline{m}}{|\overline{m+n+2}} \cdot \frac{|\overline{m+n}}{|\overline{m}\,|\overline{n}} = \frac{n(n+1)}{(m+n+1)(m+n)}$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \bar{x}^2 = \frac{n(n+1)}{(m+n+1)(m+n)} - \frac{n^2}{(m+n)^2}$$

$$= n\left[\frac{(n+1)(m+n) - n(m+n+1)}{(m+n)^2(m+n+1)}\right]$$

$$= \frac{mn}{(m+n)^2(m+n+1)}$$

$$\therefore \quad \text{s.d.} = \sqrt{\frac{mn}{m+n+1}} \cdot \frac{1}{m+n}$$

H.M. is given by

$$\frac{1}{H} = \frac{1}{\beta(m,n)} \int_0^1 x^{n-2}(1-x)^{m-1}\,dx = \frac{\beta(n-1,m)}{\beta(m,n)}$$

$$= \frac{|\overline{n-1}\,|\overline{m}}{|\overline{n+m-1}} \cdot \frac{|\overline{m+n}}{|\overline{n}\,|\overline{m}} = \frac{(m+n-1)}{n-1}$$

$$\therefore \quad H = \frac{n-1}{m+n-1}.$$

Ex. 9-18. *Prove that for the dist.*

$$dP = \frac{1}{\beta(m,n)} \frac{x^{m-1}}{(1+x)^{m+n}}\,dx, \quad 0 \le x < \infty,\ n > 2$$

variance is $\frac{m(m+n-1)}{(n-1)^2(n-2)}$. *Find also the mode and moment of rth order about the origin.*

Sol.

$$\bar{x} = \frac{1}{\beta(m,n)} \int_0^\infty \frac{x^{m+1-1}}{(1+x)^{m+n}}\,dx = \frac{\beta(m+1,n-1)}{\beta(m,n)}$$

$$= \frac{|\overline{m+1}\,|\overline{n-1}}{|\overline{m+n}} \cdot \frac{|\overline{m+n}}{|\overline{m}\,|\overline{n}} = \frac{m}{n-1}$$

$$\mu'_2(0) = \frac{1}{\beta(m,n)} \int_0^\infty \frac{x^{m+2-1}}{(1+x)^{m+n}}\,dx = \frac{\beta(m+2,n-2)}{\beta(m,n)}$$

$$= \frac{|\overline{m+2}\,|\overline{n-2}}{|\overline{m+n}} \cdot \frac{|\overline{m+n}}{|\overline{m}\,|\overline{n}} = \frac{m(m+1)}{(n-1)(n-2)}$$

Mode is that value of x for v

is maximum.

$\therefore$ Mode value x is s.t.

$f'(x) = 0$

Now $f'(x) = 0$

At $x = 0, f(x) = 0$ which is th

$\therefore$ At $x = \frac{m-1}{n+1}$, $f(x)$ is

$\therefore$ $x = \frac{m-1}{n+1}$ is the mod

Ex. 9-19. *Show that for the*

mean = variance = m. Find μ'_r

Sol.

$$\therefore \quad \mu_2 = \frac{m(m+1)}{(n-1)(n-2)} - \frac{m^2}{(n-1)^2}$$

$$= \frac{m}{(n-1)^2(n-2)}[(m+1)(n-1) - m(n-2)]$$

$$= \frac{m(m+n-1)}{(n-1)^2(n-2)}$$

$$\mu'_r(0) = \frac{1}{\beta(m,n)} \int_0^\infty \frac{x^{m+r-1}}{(1+x)^{m+n}}\, dx = \frac{\beta(m+r, n-r)}{\beta(m,n)}$$

$$= \frac{\overline{|m+r}\ \overline{|n-r}}{\overline{|m}\ \overline{|n}} = \frac{(m+r-1)(m+r-2)\ldots\ldots m}{(n-1)(n-2)\ldots(n-r)} \qquad (r < n)$$

Mode is that value of x for which

$$f(x) = \frac{1}{\beta(m,n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}}$$

is maximum.

$\therefore$ Mode value x is s.t.

$$f'(x) = 0 \text{ and } f''(x) < 0$$

Now $\qquad f'(x) = 0$ gives

$$x^{m-2}(1+x)^{m+n-1}\{(m-1)(1+x) - (m+n)x\} = 0$$

$$\therefore \quad x = 0,\ -1,\ \frac{m-1}{n+1}$$

At $x = 0, f(x) = 0$ which is the least value of $f(x)$ $x = -1$ do not belong to the range of x.

$\therefore \quad$ At $x = \dfrac{m-1}{n+1}$, $f(x)$ is maximum.

$\therefore \quad x = \dfrac{m-1}{n+1}$ is the mode. $(m > 1)$.

Ex. 9-19. *Show that for the gama dist.*

$$dF = \frac{1}{\overline{|m}} e^{-x} x^{m-1},\ 0 \le x < \infty,\ m > 0$$

mean = variance = m. Find $\mu'_r(0)$ and harmonic mean.

$$\bar{x} = \frac{1}{\overline{|m}} \int_0^\infty x^{m+1-1} e^{-x}\, dx = \frac{\overline{|m+1}}{\overline{|m}} = m.$$

Sol. $$\mu'_2(0) = \frac{1}{\overline{|m}} \int_0^\infty x^{m+2-1} e^{-x}\, dx = \frac{\overline{|m+2}}{\overline{|m}} = m(m+1)$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \bar{x}^2 = m^2 + m - m^2 = m$$

$$\mu'_r(0) = \frac{1}{\overline{|m}} \int_0^\infty x^{r+m-1} e^{-x}\, dx = \frac{\overline{|m+r}}{\overline{|m}} = m(m+1)\ldots(m+r-1)$$

H.M. is given by

$$\frac{1}{H} = \frac{1}{\Gamma(m)}\int_0^\infty x^{m-1-1}e^{-x}\,dx = \frac{\Gamma(m-1)}{\Gamma(m)} = \frac{1}{m-1}$$

$$\therefore \qquad H = m-1.$$

Ex. 9-20. *For a continuous dist.*

$$dF = y_0\, e^{-\frac{x^2}{2}}\, x^{n-1}\,dx,\ 0 \le x < \infty$$

show that $\mu'_1(0) = \sqrt{2}\,\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$

and $\qquad \mu'_2(0) = n.$

Sol. y_0 is given by

$$y_0\int_0^\infty e^{-\frac{x^2}{2}}\, x^{n-1}\,dx = 1$$

Put $\qquad \dfrac{x^2}{2} = t$

$$\therefore \qquad y_0\int_0^\infty e^{-t}.(2t)^{\frac{n}{2}-1}\,dt = 1$$

or $\qquad 2^{\frac{n}{2}-1}.y_0\int_0^\infty e^{-t}.\,t^{\frac{n}{2}-1}\,dt$

or $\qquad 2^{\frac{n}{2}-1}.y_0\,\Gamma\left(\frac{n}{2}\right) = 1.$

$$\therefore \qquad y_0 = \frac{1}{\Gamma\left(\frac{n}{2}\right).\,2^{\frac{n}{2}-1}}$$

Also be def.

$$\mu'_r(0) = y_0\int_0^\infty x^r\, e^{-\frac{x^2}{2}}.\,x^{n-1}\,dx = y_0\int_0^\infty e^{-\frac{x^2}{2}}\, x^{n+r-1}\,dx$$

Put $\qquad \dfrac{x^2}{2} = t.$

$$\therefore \qquad \mu'_r(0) = 2^{\frac{n+r-2}{2}}.\,y_0\int_0^\infty e^{-t}.\,t^{\frac{n+r}{2}-1}\,dt$$

$$= 2^{\frac{n+r-2}{2}}\,\frac{1}{2^{\frac{n-2}{2}}.\,\Gamma\left(\frac{n}{2}\right)}\cdot\Gamma\left(\frac{n+r}{2}\right) = \frac{2^{r/2}}{\Gamma\left(\frac{n}{2}\right)}\cdot\Gamma\left(\frac{n+r}{2}\right)$$

$\therefore \qquad \mu'_1(0) = \sqrt{2}$ ...

Ex. 9-21. *A frequency f" in th*...

$f(x) = \frac{1}{16}$ (

$= \frac{1}{16}$ (

$= \frac{1}{16}$ (

Find the mean and the s.d. of t...

Sol. Total frequency $= \int_{-3}^{3} f($...

$+ \frac{1}{16}$...

$= \frac{1}{8}\int_1^3$...

$= \frac{1}{8}\Big[-$...

$\bar{x} = \int_{-3}^{3} x$...

$+ \frac{1}{16}$...

For first integral change x to $-$...

$\therefore \qquad \bar{x} = -\frac{1}{16}$...

$+ \frac{1}{16}$...

$= \frac{1}{16}\int^{1}$...

$\therefore \qquad \mu_2 = \int_{-3}^{3} x^2$...

$+ \frac{1}{16}$...

$$\therefore \qquad \mu'_1(0) = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \quad \text{and} \quad \mu'_2(0) = 2\cdot\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}\right)} = n.$$

Ex. 9-21. *A frequency f^n in the range $(-3, 3)$ is defined by*

$$\begin{aligned} f(x) &= \frac{1}{16}(3+x)^2 && -3 \le x \le -1 \\ &= \frac{1}{16}(6-2x^2) && -1 \le x \le 1 \\ &= \frac{1}{16}(3-x)^2 && 1 \le x \le 3 \end{aligned}$$

Find the mean and the s.d. of the dist.

Sol. Total frequency $= \int_{-3}^{3} f(x)\,dx = \frac{1}{16}\int_{-3}^{-1}(3+x)^2\,dx$

$$+ \frac{1}{16}\int_{-1}^{1}(6-2x^2)\,dx + \frac{1}{16}\int_{1}^{3}(3-x)^2\,dx$$

$$= \frac{1}{8}\int_{1}^{3}(3-x)^2\,dx + \frac{1}{8}\int_{0}^{1}(6-2x^2)\,dx$$

$$= \frac{1}{8}\left[-\frac{(3-x)^3}{3}\right]_1^3 + \frac{1}{8}\left(6-\frac{2}{3}\right) = 1$$

$$\bar{x} = \int_{-3}^{3} x\,dx = \frac{1}{16}\int_{-3}^{-1} x(3+x)^2\,dx + \frac{1}{16}\int_{-1}^{1} x(6-2x^2)\,dx$$

$$+ \frac{1}{16}\int_{1}^{3} x(3-x)^2\,dx$$

For first integral change x to $-x$

$$\therefore \qquad \bar{x} = -\frac{1}{16}\int_{1}^{3} x(3-x)^2\,dx + \frac{1}{16}\int_{-1}^{1} x(6-2x)^2\,dx$$

$$+ \frac{1}{16}\int_{1}^{3} x(3-x)^2\,dx$$

$$= \frac{1}{16}\int_{-1}^{1} x(6-2x^2)\,dx = 0$$

$$\therefore \qquad \mu_2 = \int_{-3}^{3} x^2 f(x)\,dx = \frac{1}{16}\int_{-3}^{-1} x^2(3+x)^2\,dx + \frac{1}{16}\int_{-1}^{1} x^2(6-2x^2)\,dx$$

$$+ \frac{1}{16}\int_{1}^{3} x^2(3-x)^2\,dx$$

For first integral change x to $-x$

$$\therefore \quad \mu_2 = \frac{1}{8}\int_1^3 x^2(3-x)^2\,dx + \frac{1}{8}\int_0^1 x^2(6-2x^2)\,dx$$

$$= \frac{1}{8}\int_1^3 x^2(x^2-6x+9)\,dx + \frac{1}{8}\int_0^1 (6x^2-2x^4)\,dx$$

$$= \frac{1}{8}\left|\frac{x^5}{5}-\frac{3}{2}x^4+3x^3\right|_1^3 + \frac{1}{8}\left|2x^3-\frac{2}{5}x^5\right|_0^1$$

$$= \frac{4}{5}+\frac{1}{5} = 1$$

$\therefore$ s.d. $= 1$.

Ex. 9-22. *The dist. of a variate x in the range (0, 2) is defined by*

$$f(x) = x^3 \qquad 0 < x \le 1$$
$$= (2-x)^3 \qquad 1 < x \le 2$$

Calculate the mean, s.d. and the mean deviation about the mean of the above dist.

Sol. Total Freq. $= \int_0^2 f(x)\,dx = \int_0^1 x^3\,dx + \int_1^2 (2-x)^3\,dx$

$$= \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$$

$$\therefore \quad \mu'_1(0) = 2\int_0^2 x f(x)\,dx = 2\int_0^1 x^4\,dx + 2\int_1^2 x(2-x)^3\,dx$$

$$= \frac{2}{5} + 2\left\{\left|-\frac{(2-x)^4}{4}\cdot x\right|_1^2 + \frac{1}{4}\int_1^2 (2-x)^4\,dx\right\}$$

$$= \frac{2}{5} + 2\left\{\frac{1}{4}+\frac{1}{20}\right\} = 1$$

$$\mu'_2(0) = 2\int_0^2 x^2 f(x)\,dx = 2\int_0^1 x^5\,dx + 2\int_1^2 x^2(2-x)^3\,dx$$

$$= \frac{1}{3} + 2\left\{\left|-\frac{(2-x)^4}{4}\cdot x^2\right|_1^2 + \frac{1}{2}\int_1^2 x(2-x)^4\,dx\right\}$$

$$= \frac{1}{3}+\frac{1}{2}+\left\{\left|-\frac{(2-x)^5}{5}\cdot x\right|_1^2 + \frac{1}{5}\int_1^2 (2-x)^5\,dx\right\}$$

$$= \frac{5}{6}+\frac{1}{5}+\frac{1}{30} = \frac{16}{15}$$

$$\therefore \quad \mu_2 = \frac{16}{15} - 1 = \frac{1}{15}$$

$$\therefore \quad \text{s.d.} = \frac{1}{\sqrt{15}}$$

Mean deviation about mean

$= 2$

$= \frac{1}{10}$

$= \frac{1}{10}$

Ex. 9-23. *Find y_0, μ_2, μ_3 a*

$dF = y_0$

Sol. y_0 is given by

$y_0 \int_{-2/\alpha}^{\infty} ($

Put $1+\frac{\alpha}{2}x = y$

$\therefore \quad 1 = \frac{2y_0}{\alpha}$

Put $\frac{2}{\alpha} = \beta$

$= \beta y$

Put $y\beta^2 = t$

$= \frac{e}{\beta}$

$\therefore \quad y_0 = \frac{\beta^2}{e^\beta}$

$\mu'_r(-\beta) = y_0$

$= \frac{}{\beta}$

Put $x+\beta = \frac{y}{\beta}$

$$^2(6-2x^2)\,dx$$

$$\frac{1}{8}\int_0^1 (6x^2-2x^4)\,dx$$

$$\frac{1}{8}\left|2x^3-\frac{2}{5}x^5\right|_0^1$$

is defined by

$x \le 1$

$x \le 2$

ut the mean of the above dist.

$$2-x)^3\,dx$$

$$+\,2\int_1^2 x(2-x)^3\,dx$$

$$\frac{1}{4}\int_1^2 (2-x)^4\,dx\Bigg\}$$

$$x+2\int_1^2 x^2(2-x)^3\,dx$$

$$+\,\frac{1}{2}\int_1^2 x(2-x)^4\,dx\Bigg\}$$

$$\Bigg|_1^2+\frac{1}{5}\int_1^2 (2-x)^5\,dx\Bigg\}$$

$$\text{Mean deviation about mean} = 2\int_0^2 |x-1|\,f(x)\,dx$$

$$= 2\int_0^1 (1-x)\,x^3\,dx + 2\int_1^2 (x-1)(2-x)^3\,dx$$

$$= \frac{1}{10}+2\left\{\left|-\frac{(2-x)^4}{4}(x-1)\right|_1^2+\frac{1}{4}\int_1^2 (2-x)^4\,dx\right\}$$

$$= \frac{1}{10}+\frac{1}{10}=\frac{1}{5}.$$

Ex. 9-23. *Find y_0, μ_2, μ_3 and μ_4 for the dist.*

$$dF = y_0\left(1+\frac{\alpha}{2}x\right)^{\left(\frac{4}{\alpha^2}-1\right)} e^{-\frac{2x}{\alpha}}\,dx,\; -\frac{2}{\alpha}\le x<\infty.$$

Sol. y_0 is given by

$$y_0\int_{-2/\alpha}^{\infty}\left(1+\frac{\alpha}{2}x\right)^{\left(\frac{4}{\alpha^2}-1\right)} e^{-\frac{2x}{\alpha}}\,dx = 1$$

Put $\quad 1+\dfrac{\alpha}{2}x = y$

$$\therefore \quad 1 = \frac{2y_0}{\alpha}\int_0^{\infty} y^{\frac{4}{\alpha^2}-1}\, e^{-\frac{4}{\alpha^2}(y-1)}\,dy$$

Put $\quad \dfrac{2}{\alpha} = \beta$

$$= \beta y_0\int_0^{\infty} y^{\beta^2-1}\, e^{-\beta^2(y-1)}\,dy$$

Put $\quad y\beta^2 = t$

$$= \frac{e^{\beta^2}y_0}{\beta^{2\beta^2-1}}\int_0^{\infty} t^{\beta^2-1}\,.\,e^{-t}\,dt = \frac{e^{\beta^2}y_0}{\beta^{2\beta^2-1}}\,\Gamma(\beta^2)$$

$$\therefore \quad y_0 = \frac{\beta^{2\beta^2-1}}{e^{\beta^2}\,\Gamma(\beta^2)}$$

$$\mu'_r(-\beta) = y_0\int_{-\beta}^{\infty}(x+\beta)^r\left(1+\frac{x}{\beta}\right)^{\beta^2-1} e^{-x\beta}\,dx$$

$$= \frac{y_0}{\beta^{\beta^2-1}}\int_{-\beta}^{\infty}(x+\beta)^{\beta^2+r-1}\,e^{-x\beta}\,dx$$

Put $\quad x+\beta = \dfrac{y}{\beta}$

$$= \frac{y_0}{\beta^{\beta^2-1}} \int_0^\infty \left(\frac{y}{\beta}\right)^{\beta^2+r-1} e^{-y+\beta^2}\left(\frac{dy}{\beta}\right)$$

$$= \frac{y_0\, e^{\beta^2}}{\beta^{2\beta^2+r-1}} \int_0^\infty y^{\beta^2+r-1} e^{-y}\, dy$$

$$= \frac{1}{\Gamma(\beta^2).\beta^r} \cdot \Gamma(\beta^2+r)$$

$$\therefore \quad \mu'_1(-\beta) = \frac{\Gamma(\beta^2+1)}{\beta\,\Gamma(\beta^2)} = \beta,\ \mu'_2(-\beta) = \frac{\Gamma(\beta^2+2)}{\beta^2\,\Gamma(\beta^2)} = (\beta^2+1)$$

$$\mu'_3(-\beta) = \frac{\Gamma(\beta^2+3)}{\Gamma(\beta^2).\beta^3} = \frac{(\beta^2+2)(\beta^2+1)}{\beta}$$

and
$$\mu_4 = \frac{\Gamma(\beta^2+4)}{\Gamma(\beta^2).\beta^4} = \frac{(\beta^2+3)(\beta^2+2)(\beta^2+1)}{\beta^2}$$

$$\mu_2 = \mu'_2 - \{\mu'_1\}^2 = \beta^2+1-\beta^2 = 1$$

$$\mu_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2\{\mu'_1\}^3 = \frac{(\beta^2+2)(\beta^2+1)}{\beta} - 3\beta(\beta^2+1) + 2\beta^3$$

$$= \frac{2}{\beta} = \alpha$$

$$\mu_4 = \mu'_4 - 4\mu'_3\,\mu'_1 + 6\mu'_2\,\{\mu'_1\}^2 - 3\{\mu'_1\}^4$$

$$= \frac{(\beta^2+3)(\beta^2+2)(\beta^2+1)}{\beta^2} - 4(\beta^2+2)(\beta^2+1) + 6\beta^2(\beta^2+1) - 3\beta^4$$

$$= 3 + \frac{6}{\beta^2} = 3 + \frac{3}{2}\,\alpha^2.$$

Ex. 9-24. (*a*) *Show that for the dist.*

$$dF = y_0\left(1-\frac{x^2}{a^2}\right)^{-p} dx,\ -a \le x \le a,\ 0 < p < 1$$

$$\mu_r = \frac{(r-1)a^2}{r+1-2p}\,\mu_{r-2}$$

(*b*) *Express 'a' and 'p' in terms of σ and β_2.*

Sol. (*a*) By def.,

$$\mu_1(0) = y_0 \int_{-a}^{a} x\left(1-\frac{x^2}{a^2}\right)^{-p} dx = 0$$

$\therefore \quad \mu_r = \mu'_r$

$= y_0$

$= a^2$

$= a^2$

$+$

$= a^2$

$\therefore \quad \mu_r =$

(*b*) Put $r = 2, 4$

$\therefore \quad \sigma^2 = \mu_2$

$\therefore \quad a^2 = (3$

and $\mu_4 =$

$\therefore \quad \beta_2 =$

$\therefore \quad p =$

Ex. 9-25. *For* β_1 *(l, m) var*

$\log G =$

Sol. For β_1 (*l*, *m*) variate x,

$dF =$

$\therefore \quad \log G =$

$$\therefore \quad \mu_r = \mu'_r(0) = y_0 \int_{-a}^{a} x^r \left(1 - \frac{x^2}{a^2}\right)^{-p} dx$$

$$= y_0\, a^2 \int_{-a}^{a} x^{r-2} \left(\frac{x^2}{a^2} - 1 + 1\right)\left(1 - \frac{x^2}{a^2}\right)^{-p} dx$$

$$= a^2 . y_0 \int_{-a}^{a} x^{r-2} \left(1 - \frac{x^2}{a^2}\right)^{-p} dx$$

$$- y_0 . a^2 \int_{-a}^{a} x^{r-2} \left(1 - \frac{x^2}{a^2}\right)^{1-p} dx$$

$$= a^2 \mu_{r-2} - y_0\, a^2 \left\{ \left| \frac{x^{r-1}}{r-1} \left(1 - \frac{x^2}{a^2}\right)^{1-p} \right|_{-a}^{a} + \frac{2(1-p)}{a^2} \int_{-a}^{a} \frac{x^r}{r-1} \left(1 - \frac{x^2}{a^2}\right)^{-p} dx \right\}$$

$$= a^2 \mu_{r-2} - 2\,\frac{(1-p)}{r-1}\,\mu_r$$

$$\therefore \quad \mu_r = \frac{(r-1)a^2}{r+1-2p}\,\mu_{r-2}$$

(*b*) Put $r = 2, 4$

$$\therefore \quad \sigma^2 = \mu_2 = \frac{a^2}{3-2p}\mu_0 = \frac{a^2}{3-2p}$$

$$\therefore \quad a^2 = (3-2p)\,\sigma^2$$

and

$$\mu_4 = \frac{3a^2}{5-2p}\mu_2 = \frac{3(3-2p)}{5-2p}\,\sigma^4$$

$$\therefore \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = \frac{3(3-2p)}{5-2p}$$

$$\therefore \quad p = \frac{9-5\beta_2}{2(3-\beta_2)}.$$

Ex. 9-25. *For $\beta_1(l, m)$ variate show that*

$$\log G = \frac{\partial}{\partial l}\left\{\log\left|\overline{l} + \log\right|\overline{m} - \log\left|\overline{l+m}\right.\right\}.$$

Sol. For $\beta_1(l, m)$ variate x,

$$dF = \frac{1}{\beta(l,m)}\, x^{l-1}(1-x)^{m-1}\, dx,\ 0 \le x \le 1,\ l, m > 0$$

$$\therefore \quad \log G = \frac{1}{\beta(l,m)} \int_0^1 \log x . x^{l-1}(1-x)^{m-1}\, dx$$

$$= \frac{1}{\beta(l,m)}\frac{\partial}{\partial l}\left\{\int_0^1 x^{l-1}(1-x)^{m-1}\,dx\right\}$$

$$= \frac{1}{\beta(l,m)}\frac{\partial}{\partial l}\{\beta(l,m)\} = \frac{\partial}{\partial l}\{\log\beta\,(l,\,m)\}$$

$$= \frac{\partial}{\partial l}\left\{\log\frac{\Gamma(l)\,\Gamma(m)}{\Gamma(l+m)}\right\}$$

$$= \frac{\partial}{\partial l}\left\{\log\Gamma l+\log\Gamma m-\log\Gamma(l+m)\right\}.$$

Ex. 9-26. *Show that for the dist.*

$$dF = y_0\left\{1-\frac{|x-b|}{a}\right\}dx,\; b-a<x<b+a$$

$y_0=\dfrac{1}{a}$, *mean* $= b$ *and variance* $=\dfrac{a^2}{6}$.

Sol. y_0 is given by

$$y_0\int_{b-a}^{b+a}\left\{1-\frac{|x-b|}{a}\right\}dx=1$$

Put $\quad x-b = y$

$\therefore$
$$y_0\int_{-a}^{a}\left\{1-\frac{|y|}{a}\right\}dy=1$$

or
$$2y_0\int_{0}^{a}\left\{1-\frac{y}{a}\right\}dy=1$$

or
$$2y_0\left\{y-\frac{y^2}{2a}\right\}_0^a=1$$

$\therefore$
$$y_0 = \frac{1}{a}$$

$$\text{Mean} = y_0\int_{b-a}^{b+a}x\left\{1-\frac{|x-b|}{a}\right\}dx$$

$$= y_0\int_{-a}^{a}(y+b)\left\{1-\frac{|y|}{a}\right\}dy=2by_0\int_0^a\left(1-\frac{y}{a}\right)dy$$

$$= by_0\,a=b$$

$$\mu_2 = y_0\int_{b-a}^{b+a}(x-b)^2\left\{1-\frac{|x-b|}{a}\right\}dx=y_0\int_{-a}^{a}y^2\left(1-\frac{|y|}{a}\right)dy$$

$$= 2y_0\int_0^a y^2\left(1-\frac{y}{a}\right)dy = 2y_0\frac{a^3}{12}=\frac{a^2}{6}.$$

Ex. 9-27. *Show that the f"*

$F(x) = \{0 \;\; \frac{1}{2}(\;\; 1$

is a distribution function.

Sol. Evidently

$F(x =$

and $\quad F'(x) = \{0 \;\; \frac{1}{2a} \;\; 0$

Evidently $F'(x)\geq 0$

$\therefore\;$ $F(x)$ is a distribution functi

Ex. 9-28. *For the distribution*

$f(x) = \frac{1}{\pi}\,\frac{\;}{1+}$

Find mean, mode, median, vari

Sol. Distribution function is gi

$F(x) = \frac{1}{\pi}\int_{-\infty}^{x}$

$= \frac{1}{\pi}\tan$

For median, $\quad F(x) = \frac{1}{2}$

$\therefore \quad x = 0$

For first quartile, $F(x) = \frac{1}{4}$

or $\quad \frac{1}{\pi}\tan^{-1}x+\frac{1}{2} = \frac{1}{4}$

or $\quad \tan^{-1}x = -\frac{\pi}{4}$

$\therefore \quad x = -1$

For third quartile, $F(x) = \frac{3}{4}$

$\therefore \quad x = 1$

$\text{Mean} = \frac{1}{\pi}\int_{-\infty}^{\infty}$

$\therefore \quad \mu_2 = \frac{1}{\pi}\int_{-\infty}^{\infty}$

$= \frac{2}{\pi}\{|x$

Ex. 9-27. *Show that the f*^{n}

$$F(x) = \begin{cases} 0 & x < -a \\ \dfrac{1}{2}\left(\dfrac{x}{a}+1\right) & -a \le x \le a \\ 1 & x > a \end{cases}$$

is a distribution function.

Sol. Evidently

$$F(x=\infty) = 1, \quad F(x=-\infty) = 0$$

and
$$F'(x) = \begin{cases} 0 & x < -a \\ \dfrac{1}{2a} & -a \le x \le a \\ 0 & x > a \end{cases}$$

Evidently $F'(x) \ge 0$

$\therefore$ $F(x)$ is a distribution function.

Ex. 9-28. *For the distribution with density f*^{n}.

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2} \quad -\infty < x < \infty$$

Find mean, mode, median, variance, first and third quartiles and distribution function.

Sol. Distribution function is given by

$$F(x) = \frac{1}{\pi}\int_{-\infty}^{x}\frac{dx}{1+x^2} = \frac{1}{\pi}\left\{\tan^{-1}x + \frac{\pi}{2}\right\}$$

$$= \frac{1}{\pi}\tan^{-1}x + \frac{1}{2}$$

For median, $\quad F(x) = \dfrac{1}{2}$

$\therefore \quad x = 0$

For first quartile, $F(x) = \dfrac{1}{4}$

or
$$\frac{1}{\pi}\tan^{-1}x + \frac{1}{2} = \frac{1}{4}$$

or
$$\tan^{-1}x = -\frac{\pi}{4}$$

$\therefore \quad x = -1$

For third quartile, $F(x) = \dfrac{3}{4}$

$\therefore \quad x = 1$

$$\text{Mean} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x}{1+x^2}\,dx = 0$$

$\therefore$
$$\mu_2 = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x^2}{1+x^2}\,dx = \frac{2}{\pi}\int_0^{\infty}\left\{1 - \frac{1}{1+x^2}\right\}dx$$

$$= \frac{2}{\pi}\left\{|x|_0^{\infty} - \frac{\pi}{2}\right\}$$

Evidently μ_2 does not exist.

For modal value $x, f'(x) = 0$

$$\therefore \quad \frac{1}{\pi}\frac{-2x}{(1+x^2)^2} = 0$$

$\therefore$ Modal value $= 0$.

Ex. 9-29. *For the distribution given by*

$$f(x) = \begin{cases} \dfrac{2(b+x)}{b(a+b)}, & -b \le x \le 0 \\ \dfrac{2(a-x)}{a(a+b)}, & 0 \le x \le a \end{cases}$$

find mean, median and variance.

Sol. Total prob. $= \int_{-b}^{a} f(x)\,dx$

$$= \frac{2}{b(a+b)}\int_{-b}^{0}(b+x)\,dx + \frac{2}{a(a+b)}\int_{0}^{a}(a-x)\,dx$$

$$= \frac{2}{b(a+b)}\left\{b^2 - \frac{b^2}{2}\right\} + \frac{2}{a(a+b)}\left\{a^2 - \frac{a^2}{2}\right\}$$

$$= 1$$

$$\text{mean} = E(x) = \int_{-b}^{a} x f(x)\,dx$$

$$= \frac{2}{b(a+b)}\int_{-b}^{0} x(b+x)\,dx + \frac{2}{a(a+b)}\int_{0}^{a} x(a-x)\,dx$$

$$= \frac{2}{b(a+b)}\left\{\frac{-b^3}{6}\right\} + \frac{2}{a(a+b)}\left\{\frac{a^3}{6}\right\} = \frac{a-b}{3}$$

$$\mu'_2(0) = \int_{-b}^{a} x^2 f(x)\,dx$$

$$= \frac{2}{b(a+b)}\int_{-b}^{0} x^2(b+x)\,dx + \frac{2}{a(a+b)}\int_{0}^{a} x^2(a-x)\,dx$$

$$= \frac{2}{b(a+b)}\left\{\frac{b^4}{12}\right\} + \frac{2}{a(a+b)}\left\{\frac{a^4}{12}\right\}$$

$$= \frac{b^3+a^3}{6(a+b)} = \frac{a^2+b^2-ab}{6}$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \bar{x}^2$$

$$= \frac{a^2+b^2-ab}{6} - \left(\frac{a-b}{3}\right)^2$$

To find median.

Let M be the median

Then $\frac{1}{2} = \int$

Now $\int_{-b}^{0} f(x)\,dx = $

Now $\frac{b}{a+b} \lesseqgtr \frac{1}{2}$ iff b

$\Rightarrow \int_{-b}^{0} f(x)\,dx \lesseqgtr \frac{1}{2}$ iff

$\therefore$ $M \gtreqless 0$ $b \lesseqgtr$

Let a > b Than

$\therefore$ (1) $\Rightarrow \frac{1}{2} = \int_{-b}^{0} f(x)\,dx$

$\Rightarrow \frac{a-b}{2(a+b)} =$

$\Rightarrow$ $2M^2$ -

$\Rightarrow$ $M = a$

Since M is to be less than a,

$\therefore$ $M = a$ -

let $a < b$

Here $M < 0$

$\therefore$ (1) $\Rightarrow \frac{1}{2} =$

$\Rightarrow$ $b(a+b) = 4$

$\Rightarrow$ $= 2($

$\Rightarrow$ $M = -$

$$\frac{2}{a+b)}\int_0^a (a-x)\,dx$$

$$\frac{2}{a+b)}\left\{a^2-\frac{a^2}{2}\right\}$$

$$\frac{2}{a(a+b)}\int_0^a x(a-x)\,dx$$

$$\left\{\frac{a^3}{6}\right\}=\frac{a-b}{3}$$

$$\frac{2}{a(a+b)}\int_0^a x^2(a-x)\,dx$$

$$\left\{\frac{a^4}{12}\right\}$$

$$= \frac{a^2+b^2+ab}{18}$$

To find median.

Let M be the median

Then $$\frac{1}{2} = \int_{-b}^{M} f(x)\,dx. \qquad ...(1)$$

Now $$\int_{-b}^{0} f(x)\,dx = \frac{b}{a+b}$$

Now $$\frac{b}{a+b} \lesseqgtr \frac{1}{2} \quad \text{iff } b \lesseqgtr a.$$

$$\Rightarrow \int_{-b}^{0} f(x)\,dx \lesseqgtr \frac{1}{2} \quad \text{iff } b \lesseqgtr a.$$

$$\therefore \quad M \gtreqless 0 \qquad b \lesseqgtr a. \qquad ...(2)$$

Let a > b Than $M > 0$

$$\therefore \quad (1) \Rightarrow \frac{1}{2} = \int_{-b}^{0} f(x)\,dx + \int_{0}^{M} f(x)\,dx$$

$$= \frac{b}{a+b} + \frac{2}{a(a+b)}\int_0^M (a-x)\,dx$$

$$\Rightarrow \quad \frac{a-b}{2(a+b)} = \frac{2}{a(a+b)}\left[aM-\frac{M^2}{2}\right]$$

$$\Rightarrow \quad 2M^2-4aM+a(a-b)=0$$

$$\Rightarrow \quad M = a \pm \sqrt{\frac{a^2+ab}{2}}$$

Since M is to be less than a, we take negative sign only.

$$\therefore \quad M = a-\sqrt{\frac{a(a+b)}{2}}$$

let *a* < *b*

Here $M < 0$

$$\therefore \quad (1) \Rightarrow \frac{1}{2} = \frac{2}{b(a+b)}\int_{-b}^{M} (b+x)\,dx$$

$$\Rightarrow \quad b(a+b) = 4\left\{b(M+b)+\frac{M^2-b^2}{2}\right\}$$

$$\Rightarrow \quad = 2(M+b)^2$$

$$\Rightarrow \quad M = -b \pm \sqrt{\frac{b(a+b)}{2}}$$

$$= -b + \sqrt{\frac{b(a+b)}{2}}$$

{We neglect – sign as we must have $M > -b$}

If $a = b$ since we have

$$\int_{-b}^{0} f(x)\,dx = \frac{b}{a+b} = \frac{1}{2}$$

0 is the median

Ex. 9-30. *Let* $f(x) = ke^{-\alpha x}(1 - e^{-\alpha x}), x > 0, \alpha\ 0$

(*i*) *Find k such that f(x) is a density* f^n.

(*ii*) *Find the corresponding cumulative distribution* f^n.

(*iii*) *Find* $P(x > 1)$.

Sol. (*i*) k is given by

$$k\int_0^\infty e^{-\alpha x}(1-e^{-\alpha x})\,dx = 1$$

$$\Rightarrow \qquad = k\left[\frac{e^{-\alpha x}}{-\alpha} - \frac{e^{-2\alpha x}}{-2\alpha}\right]_0^\infty$$

$$= k\left[\frac{1}{\alpha} - \frac{1}{2\alpha}\right] = \frac{k}{2\alpha}$$

$$\Rightarrow \qquad k = 2\alpha$$

$$\therefore \qquad f(x) = 2\alpha e^{-\alpha x}(1 - e^{-\alpha x})$$

(*ii*) Cumulative distribution function is given by

$$F(x) = \int_0^x f(x)\,dx$$

$$= 2\alpha\int_0^x e^{-\alpha x}(1-e^{-\alpha x})\,dx$$

$$= 2\alpha\left[\frac{e^{-\alpha x}}{-\alpha} - \frac{e^{-2\alpha x}}{-2\alpha}\right]_0^x$$

$$= 2\alpha\left[-\frac{1}{\alpha}(e^{-\alpha x} - 1) + \frac{1}{2\alpha}(e^{-2\alpha x} - 1)\right]$$

$$= 1 - 2e^{-\alpha x} + e^{-2\alpha x}$$

(*iii*)
$$P(x > 1) = 1 - P(x \le 1)$$
$$= 1 - F(1)$$
$$= 1 - \{1 - 2e^{-\alpha} + e^{-2\alpha}\}$$
$$= 2e^{-\alpha} - e^{-2\alpha}.$$

Ex. 9-31. *The prob. density function of coded measurements of pitch diameter of threads of a fitting is given by*

$$f(x) = \frac{1}{(1+x)^2},\ 0 \le x < \infty$$

Find the distribution functio[n] ... the quartiles of the distribution. I... exist.

Sol. Distribution function is

$$F(x) = \int_0^x$$

$$= 1 -$$

and $F(x) = 0$ f

(*i*) $\therefore \quad P(x > 2) = 1 -$

(*ii*) For median value x, $F($

or $\dfrac{x}{1+x} = \dfrac{1}{2}$

$\therefore \quad x = 1$

For first quartile value x, $F(x$

or $\dfrac{x}{1+x} = \dfrac{1}{4}$

$\therefore \quad x = \dfrac{1}{3}$

And for third quartile value x

$\therefore \quad \dfrac{x}{1+x} = \dfrac{3}{4}$

$\therefore \quad x = 3$

Mean $= \displaystyle\int_0^\infty \frac{x}{(1+x)^2}\,dx = \int_0^\infty$

$= |\log$

Since as $x \to \infty$, $\log(1+x) \to$

Ex. 9-32. *A bombing plane c... If a bomb falls within 40 feet of t... traffic. With a certain bomb-sigh... function*

$$f(x) = \frac{10\ldots}{1\ldots}$$

$$= \frac{10\ldots}{1\ldots}$$

$$= 0$$

$n f^n$.

$(e^{-2\alpha x}-1)\big]$

rements of pitch diameter of threads

Find the distribution function of the dist. Hence obtain (i) $P(x>2)$, (ii) the median and the quartiles of the distribution. Investigate whether the mean and the variance of the dist. exist.

Sol. Distribution function is given by

$$F(x) = \int_0^x \frac{1}{(1+x)^2}\,dx = \left| -\frac{1}{1+x} \right|_0^x$$

$$= 1-\frac{1}{1+x} = \frac{x}{1+x} \quad \text{for } x \geq 0$$

and $\quad F(x) = 0 \;\text{ for }\; x<0$

(i) $\therefore \quad P(x>2) = 1-P(x\leq 2) = 1-F(2) = 1-\dfrac{2}{1+2} = \dfrac{1}{3}$

(ii) For median value x, $F(x) = \dfrac{1}{2}$

or $$\frac{x}{1+x} = \frac{1}{2}$$

$\therefore \quad x = 1$

For first quartile value x, $F(x) = \dfrac{1}{4}$

or $$\frac{x}{1+x} = \frac{1}{4}$$

$\therefore \quad x = \dfrac{1}{3}$

And for third quartile value x, $F(x) = \dfrac{3}{4}$

$\therefore$ $$\frac{x}{1+x} = \frac{3}{4}$$

$\therefore \quad x = 3$

$$\text{Mean} = \int_0^\infty \frac{x}{(1+x)^2}\,dx = \int_0^\infty \frac{1}{1+x}\,dx - \int_0^\infty \frac{dx}{(1+x)^2}$$

$$= \left|\log(1+x)\right|_0^\infty - 1$$

Since as $x \to \infty$, $\log(1+x) \to \infty$, mean does not exist. Hence variance will also not exit.

Ex. 9-32. *A bombing plane carrying three bombs flies directly above a railroad track. If a bomb falls within 40 feet of track, the track will be sufficiently damaged to disrupt the traffic. With a certain bomb-sight the points of impact of a bomb have the prob. density function*

$$f(x) = \frac{100+x}{10^4} \qquad \textit{when } -100 \leq x \leq 0$$

$$= \frac{100-x}{10^4} \qquad \textit{when } 0 \leq x \leq 100$$

$$= 0 \qquad \textit{elsewhere}$$

where x represents the vertical deviation from the aiming point, which is the track in this case. Find the distribution function. If all the three bombs are used, what is the prob. that the track will be damaged ?

Sol. Let $F(x)$ be distribution function.

Then $$F(x) = \int_{-100}^{x} f(x)\,dx = \int_{-100}^{x} \frac{100+x}{10^4}\,dx, \text{ if } -100 \le x \le 0$$

$$= \left[\frac{100x+(x^2/2)}{10^4}\right]_{-100}^{x} = \frac{1}{10^4}\left[100x + \frac{x^2}{2} + \frac{10^4}{2}\right]$$

$$F(x) = \int_{-100}^{0} \frac{100+x}{10^4}\,dx + \int_{0}^{x} \frac{100-x}{10^4}, \text{ if } 0 \le x \le 100$$

$$= \frac{1}{10^4}\left[100x + \frac{x^2}{2}\right]_{-100}^{0} + \frac{1}{10^4}\left[100x - \frac{x^2}{2}\right]_{0}^{x}$$

$$= \frac{1}{10^4}\left[100x - \frac{x^2}{2} + \frac{10^4}{2}\right]$$

$F(x) = 0$ if $x < -100$

and $F(x) = 1$ if $x > 100$

Now prob. for a bomb to fall within 40 feet of the track

$$= \int_{-40}^{0} f(x)dx + \int_{0}^{40} f(x)\,dx$$

$$= \int_{-40}^{0} \frac{100+x}{10^4}\,dx + \int_{0}^{40} \frac{100-x}{10^4}\,dx$$

$$= \frac{2}{10^4}\left[100x - \frac{x^2}{2}\right]_{0}^{40} = \frac{2}{10^4}\{4000-800\} = \frac{16}{25}$$

$\therefore$ Prob. for a bomb not to fall within 40 feet of the track

$$= 1 - \frac{16}{25} = \frac{9}{25}$$

$\therefore$ Prob. for all the three bombs not to fall within 40 feet of the track

$$= \left(\frac{9}{25}\right)^3$$

$\therefore$ Prob. of at least one bomb falling within 40 feet of the track

$$= 1 - \left(\frac{9}{25}\right)^3$$

$\therefore$ Prob. of the track being damaged = Prob. of at least one bomb falling within 40 feet of the track

$$= 1 - \left(\frac{9}{25}\right)^3.$$

Ex. 9-33. *Suppose the life in function*

$$f(x) = \frac{100}{x^2}$$
$$= 0$$

Find the prob. that none of th replaced during the first 150 hour original tubes will have been repla

Sol. Let x hours be the life of t

Then $P\{x \le 150\} = \int_0^{150} f($

$\therefore$ By compound prob. theorer during the first 150 hours of operat

$$= \left(\frac{1}{3}\right)^3$$

Also $P(x > 150) = 1 - \frac{1}{3}$

$\therefore$ Prob. that none of the three operation

$$= \left(\frac{2}{3}\right)^3$$

Ex. 9-34. *Assuming $F(x)$ to be c x and given that $F(x_1) = 0{\cdot}5$, $F(x_2) =$ lie between x_1 and x_2; x_2 and x_3. Ca and X_2 will lie between $-\infty$ to x_1 a $-\infty$ to ∞.*

Sol. Let $f(x)$ be the density fun

Then $P\{x_1 \le x \le x_2\} = \int_{x_1}^{x_2} f($

$= \int_{x_1}^{-\infty} f($

$= \int_{-\infty}^{x_2} f($

$= F(x_2)$

$P\{x_2 \le x \le x_3\} = \int_{x_2}^{x_3} f($

$= F(x_3)$

Now $P\{-\infty < X_1 < x_1\} = F($

$P\{x_3 \le X_2 < \infty\} = 1 - F$

$= 1 - F$

point, which is the track in this
s are used, what is the prob. that

dx, if $-100 \le x \le 0$

$\left[100x + \frac{x^2}{2} + \frac{10^4}{2}\right]$

$\frac{x}{}$, if $0 \le x \le 100$

$\frac{1}{0^4}\left[100x - \frac{x^2}{2}\right]_0^x$

ack

$\frac{-x}{4}\,dx$

$\{4000-800\} = \frac{16}{25}$

track

) feet of the track

of the track

ast one bomb falling within 40 feet

Ex. 9-33. *Suppose the life in hours of a certain kind of ratio tube has the density function*

$$
\begin{aligned}
f(x) &= \frac{100}{x^2} && \text{when } x \ge 100 \\
&= 0 && \text{otherwise}
\end{aligned}
$$

Find the prob. that none of the three such tubes in a given radioset will have to be replaced during the first 150 *hours of operation? What is the prob. that all three of the original tubes will have been replaced during the first* 150 *hours ?*

Sol. Let x hours be the life of the tube.

Then $$P\{x \le 150\} = \int_0^{150} f(x)\,dx = 100 \int_{100}^{150} \frac{1}{x^2}\,dx = \frac{1}{3}$$

$\therefore$ By compound prob. theorem, prob. that all the three tubes will have to be replaced during the first 150 hours of operation.

$$= \left(\frac{1}{3}\right)^3 = \frac{1}{27}$$

Also $$P(x > 150) = 1 - \frac{1}{3} = \frac{2}{3}$$

$\therefore$ Prob. that none of the three tubes will have to be replaced during first 150 hours of operation

$$= \left(\frac{2}{3}\right)^3 = \frac{8}{27}.$$

Ex. 9-34. *Assuming $F(x)$ to be cumulative probability distribution function for a variable x and given that $F(x_1) = 0{\cdot}5$, $F(x_2) = 0{\cdot}7$ and $F(x_3) = 0{\cdot}8$, find the prob. that the variable will lie between x_1 and x_2; x_2 and x_3. Calculate the prob. that two independent observations X_1 and X_2 will lie between $-\infty$ to x_1 and x_3 to ∞. It may be assumed that x takes values from $-\infty$ to ∞.*

Sol. Let $f(x)$ be the density function.

$$
\begin{aligned}
\text{Then } P\{x_1 \le x \le x_2\} &= \int_{x_1}^{x_2} f(x)dx \\
&= \int_{x_1}^{-\infty} f(x)dx + \int_{-\infty}^{x_2} f(x)\,dx \\
&= \int_{-\infty}^{x_2} f(x)dx - \int_{-\infty}^{x_1} f(x)\,dx \\
&= F(x_2) - F(x_1) = 0{\cdot}7 - 0{\cdot}5 = 0{\cdot}2 \\
P\{x_2 \le x \le x_3\} &= \int_{x_2}^{x_3} f(x)dx = \int_{-\infty}^{x_3} f(x)\,dx - \int_{-\infty}^{x_2} f(x)\,dx \\
&= F(x_3) - F(x_2) = 0{\cdot}8 - 0{\cdot}7 = 0{\cdot}1 \\
\text{Now } P\{-\infty < X_1 < x_1\} &= F(x_1) = 0{\cdot}5 \\
P\{x_3 \le X_2 < \infty\} &= 1 - F\{-\infty < X_2 < x_3\} \\
&= 1 - F(x_3) = 1 - 0{\cdot}8 = 0{\cdot}2
\end{aligned}
$$

Since X_1 and X_2 are independent observations, reqd. prob.

$$= (0{\cdot}5)\,(0{\cdot}2) = 0{\cdot}1.$$

Ex. 9-35. *A distribution function is defined as follows*

$$F(x) = \begin{cases} 0 & x \le 1 \\ \frac{1}{16}(x-1)^4 & 1 \le x \le 3 \\ 1 & x > 3 \end{cases}$$

Find the density function f(x). Find the mean of x and the median.

Sol. Density function is given by

$$f(x) = F'(x) = \frac{1}{4}(x-1)^3 \quad 1 \le x \le 3$$
$$= 0 \qquad \text{otherwise}$$

$$\therefore \quad \text{Mean} = \frac{1}{4}\int_1^3 x(x-1)^3\,dx = \frac{1}{4}\left[\left|\frac{(x-1)^4}{4}x\right|_1^3 - \frac{1}{4}\int_1^3 (x-1)^4\,dx\right]$$
$$= 3 - \frac{1}{80}\left|(x-1)^5\right|_1^3 = 3 - \frac{2}{5} = 2{\cdot}6$$

For median $\quad F(x) = \frac{1}{2}$

$$\therefore \quad \frac{1}{16}(x-1)^4 = \frac{1}{2}$$

$$\therefore \quad x = 1 + (8)^{\frac{1}{4}}.$$

Ex. 9-36. *Determine m so that the following function represents the density function*

$$f(x) = \begin{cases} 0 & x \le -1 \\ m(x+1) & -1 < x \le 3 \\ 4m & 3 < x \le 4 \\ 0 & x > 4 \end{cases}$$

Find the value of x about which the mean deviation of this dist is least.

Sol. m is given by

$$\int_{-1}^{4} f(x)dx = 1$$

or $$\int_{-1}^{3} m(x+1)\,dx + \int_{3}^{4} 4m\,dx = 1$$

or $$m\left|\frac{(x+1)^2}{2}\right|_{-1}^{3} + 4m|x|_3^4 = 1$$

or $$8m + 4m = 1$$

or $$m = \frac{1}{12}$$

Since the mean deviation is least about median, it is required to find median. Let it be 'a'

Then $\int_{-1}^{a} f(x)dx = \frac{1}{2}.$

Let if possible $a > 3$

Then $m\int_{-1}^{3}$ (

or $8m + 4m\,(a-3) = \frac{1}{2}$

or $\frac{1}{3}(a-3) = \frac{1}{2} -$

which is not possible

$\therefore \quad a \not> 3$ i.e., $a < 3$

$\Rightarrow \quad (a+1)^2 = 12$

$\therefore \quad a = 2\sqrt{3}$

Ex. 9-37. *A country filling st*
volume x of sales in thousands of
must be the capacity of its tank in
given week shall be 0·01?

Sol. Let V be the volume of c
Then prob. that the supply wi
$= P(x$

$\therefore \quad P(x \ge V) = 0{\cdot}01$

$\therefore \quad P(x < V) = 0{\cdot}9$

$\therefore \quad 5\int_0^V (1-x)^4 dx = 0{\cdot}9$

or $1 - (1-V)^5 = 0{\cdot}9$

or $(1-V)^5 = 0{\cdot}01$

$\therefore \quad V = 0{\cdot}6$

$\therefore$ Tank capacity $= 602$

Ex. 9-38. *For continuous vari*
from the median.

Sol. Let x be a continuous **va**
Now by def., **mean deviation**

$F(a) = E\,|x-a| = \int_{-\infty}^{\infty}$

$= \int_{-\infty}^{a}$

Differentiating *w.r.t.* 'a' unde

$F'(a) = \int_{-\infty}^{a}$

and $F''(a) = f(a)$

Then $$\int_{-1}^{a} f(x)dx = \frac{1}{2}.$$

Let if possible $a > 3$

Then $$m\int_{-1}^{3} (x+1)dx + 4m\int_{3}^{a} dx = \frac{1}{2}$$

or $$8m + 4m(a-3) = \frac{1}{2}$$

or $$\frac{1}{3}(a-3) = \frac{1}{2} - 8m = \frac{1}{2} - \frac{2}{3} = -\frac{1}{6}$$

which is not possible

$$\therefore \quad a \ngtr 3 \text{ i.e., } a < 3 \qquad \therefore \quad m\int_{-1}^{a} (x+1)dx = \frac{1}{2}$$

$$\Rightarrow \quad (a+1)^2 = 12$$

$$\therefore \quad a = 2\sqrt{3} - 1.$$

Ex. 9-37. *A country filling station is supplied with gasoline once a week. If its weekly volume x of sales in thousands of gallons is distributed by* $f(x) = 5(1-x)^4$, $0 < x < 1$, *what must be the capacity of its tank in order that the prob. that its supply will be exhausted in a given week shall be* 0·01?

Sol. Let V be the volume of capacity of the tank in thousands of gallons.

Then prob. that the supply will be exhausted in a given week

$$= P(x \geq V)$$

$$\therefore \quad P(x \geq V) = 0{\cdot}01$$

$$\therefore \quad P(x < V) = 0{\cdot}99$$

$$\therefore \quad 5\int_{0}^{V} (1-x)^4 dx = 0{\cdot}99$$

or $$1 - (1-V)^5 = 0{\cdot}99$$

or $$(1-V)^5 = 0{\cdot}01$$

$$\therefore \quad V = 0{\cdot}602$$

$$\therefore \quad \text{Tank capacity} = 602 \text{ gallons.}$$

Ex. 9-38. *For continuous variable, show that the mean deviation is least when measured from the median.*

Sol. Let x be a continuous variable with density function $f(x)$.

Now by def., mean deviation about an arbitrary point 'a' is given by

$$F(a) = E|x-a| = \int_{-\infty}^{\infty} |x-a| f(x)\, dx$$

$$= \int_{-\infty}^{a} (a-x) f(x)\, dx + \int_{a}^{\infty} (x-a) f(x)\, dx$$

Differentiating *w.r.t.* 'a' under the sign of integration

$$F'(a) = \int_{-\infty}^{a} f(x)dx - \int_{a}^{\infty} f(x)\, dx$$

and $$F''(a) = f(a) + f(a) = 2f(a)$$

For $F(a)$ to be minimum, 'a' is given by

$$F'(a) = 0 \text{ i.e., } \int_{-\infty}^{a} f(x)\,dx = \int_{a}^{\infty} f(x)\,dx$$

which implies that 'a' is the median.

If $f(a) \neq 0$. $F''(a) > 0$ $\quad (\because f(a) > 0)$

$\therefore$ $F(a)$ is minimum when 'a' is median.

If $f(a) = 0$. $f(x)$ is minimum for $x = a$ $\quad (\because f(x) \not< 0)$.

First derivative of $f(x)$ which is not zero for $x = a$ is of even order and is positive.

$\therefore$ First derivative of $F(a)$ which is not zero when 'a' is the median of even order and is positive.

$\therefore$ $F(a)$ is minimum when a is median.

9.2. Tchebycheff's Inequality

Let x be a continuous variable with density function $f(x)$ and expected value zero. The variance of x is given by

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{-\infty}^{-\sigma_x.t} x^2 f(x)\,dx + \int_{-\sigma_x.t}^{\sigma_x.t} x^2 f(x)\,dx + \int_{\sigma_x.t}^{\infty} x^2 f(x)\,dx$$

Now $$\int_{-\sigma_x.t}^{\sigma_x.t} x^2 f(x)\,dx \geq 0$$

$$\therefore \quad \sigma_x^2 \geq \int_{-\infty}^{-\sigma_x.t} x^2 f(x)\,dx + \int_{\sigma_x.t}^{\infty} x^2 f(x)\,dx$$

$$\geq \sigma_x^2 t^2 \left\{ \int_{-\infty}^{-\sigma_x.t} f(x)\,dx + \int_{\sigma_x.t}^{\infty} f(x)\,dx \right\}$$

$$= \sigma_x^2 t^2 P\{|x| \geq \sigma_x t\}$$

$$\therefore \quad \frac{1}{t^2} \geq P\{|x| \geq \sigma_x t\}$$

Put $$x = y - \bar{y}$$

Then $$E(x) = 0 \text{ and } \sigma_x^2 = E(y - \bar{y})^2 = \sigma_y^2$$

$$\therefore \quad \frac{1}{t^2} \geq P\{|y - \bar{y}| \geq \sigma_y.t\}$$

which is Tchebycheff's inequality.

9.3. State and prove weak law of large numbers

Let $x_1, x_2, \ldots\ldots x_n$ be n independent random variables distributed in the same form with mean m and $\bar{x} = \dfrac{x_1 + x_2 + \ldots\ldots + x_n}{n}$.

Then for any fixed $\varepsilon > 0$,

$$\operatorname*{Lt}_{n \to \infty} P\{|\bar{x} - m| > \varepsilon\} = 0$$

Assume that the variances of $x_1, x_2, \ldots\ldots x_n$ exist and are equal to σ^2

Then $\quad \operatorname{Var}(\bar{x}) = E\{$

$= \frac{1}{n^2}$

$= \frac{1}{n^2}$

$= \frac{n\sigma}{n^2}$

$\therefore$ From Tchebycheff's inequ

$\frac{1}{t^2} \geq P[$

Let $\quad t\frac{\sigma}{\sqrt{n}} > \varepsilon$

$\therefore \quad t > \frac{\sqrt{n}}{\sigma}$

$\therefore \quad \frac{\sigma^2}{n\varepsilon^2} > P\{|$

which implies that for given $\varepsilon > 0$,

$\therefore \quad \operatorname*{Lt}_{n\to\infty} P\{|\bar{x}$ -

Note. (1) The above law rem

(2) The weak law of large num

In fact, given any positive numbe

$P\{|\bar{x} - m| \geq \varepsilon\} \leq \delta, \quad n \geq$

The weak law states that $|\bar{x}$ -

might be that for some n it was lar

strong law says that the prob. of

variables which are identically d

cases further conditions must be

Ex. 9-39. *Define stochastic*

series of Bernoullian trials, the

prob. of success in each trial as

Sol. Def. A variate x_n is sa

parameter θ if given any positive

$P\{|x_n - \theta| \geq \varepsilon\} < \delta$ for

Let $x_1, x_2, \ldots. x_n$ be the variat

$x_i = 1$ i

$= 0$ i

Then number of successes is

$m = x_1$

and $\quad E(x_i) = 1.$

$\therefore \quad \operatorname{Var}(x_i) = p$ -

'x

$\because\ f(x) \not< 0$).

f even order and is positive.

s the median of even order and is

(x) and expected value zero. The

$x)\, dx + \int_{\sigma_x \cdot t}^{\infty} x^2 f(x)\, dx$

$x)\, dx$

$f(x)\, dx \Big\}$

$_y^2$

distributed in the same form with

are equal to σ^2

Then
$$\text{Var}(\bar{x}) = E\left\{\frac{x_1 + x_2 + \ldots . x_n}{n} - m\right\}^2$$
$$= \frac{1}{n^2} E\{(x_1 - m) + (x_2 - m) + \ldots\ldots.. + (x_n - m)\}^2$$
$$= \frac{1}{n^2} E\{(x_1 - m)^2 + E(x_2 - m)^2 + \ldots\ldots + E(x_n - m)^2\}$$
$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\therefore$ From Tchebycheff's inequality
$$\frac{1}{t^2} \geq P\left[|\bar{x} - m| \geq \frac{\sigma}{\sqrt{n}} t\right]$$

Let
$$t\frac{\sigma}{\sqrt{n}} > \varepsilon$$
$$\therefore \quad t > \frac{\sqrt{n}\,\varepsilon}{\sigma}$$
$$\therefore \quad \frac{\sigma^2}{n\varepsilon^2} > P\{|\bar{x} - m| > \varepsilon\}$$

which implies that for given $\varepsilon > 0$, the prob. can be made as small as we please by increasing n.
$$\therefore \quad \underset{n \to \infty}{\text{Lt}}\ P\{|\bar{x} - m| > \varepsilon\} = 0$$

Note. (1) The above law remains true even if we discard the requirement that σ^2 exists.

(2) The weak law of large numbers states a limiting property of sums of random variables. In fact, given any positive numbers ε and δ there is an n *s.t.*
$$P\{|\bar{x} - m| \geq \varepsilon\} \leq \delta, \quad n \geq N$$

The weak law states that $|\bar{x} - m|$ is ultimately small but not that every value is small; it might be that for some n it was large, although such cases could only occur infrequently. The strong law says that the prob. of such happening is extremely small. The law is true for variables which are identically distributed under the sole condition that μ exists; in other cases further conditions must be added.

Ex. 9-39. *Define stochastic convergence of the variate and show that in an infinite series of Bernoullian trials, the proportion of successes converges stochastically to the prob. of success in each trial as the number of trials increases indefinitely.*

Sol. Def. A variate x_n is said to converge stochastically (or in probability sense) to parameter θ if given any positive numbers ε and δ there is N *s.t.*
$$P\{|x_n - \theta| \geq \varepsilon\} < \delta \text{ for } n > N$$

Let $x_1, x_2, \ldots . x_n$ be the variates
$$x_i = 1 \text{ if } i\text{th trial results in success}$$
$$= 0 \text{ if } i\text{th trial results in failure}$$

Then number of successes is given by
$$m = x_1 + x_2 + \ldots . + x_n$$

and
$$E(x_i) = 1 . p + 0\,(1 - p) = p,\ E(x_i^2) = p$$
$$\therefore \quad \text{Var}(x_i) = p - p^2 = pq$$

where p is the prob. of success in each trial

$$\therefore \quad E(m) = np \text{ and } \operatorname{Var}(m) = nqp$$

$$\therefore \quad E\left(\frac{m}{n}\right) = p$$

Also $$\operatorname{Var}\left(\frac{m}{n}\right) = \frac{1}{n^2} \operatorname{Var}(m) = \frac{pq}{n}$$

$\therefore$ From Tchebycheff's inequality

$$P\left\{\left|\frac{m}{n} - p\right| \geq \sqrt{\frac{pq}{n}}\, t\right\} \leq \frac{1}{t^2}$$

Let $$\sqrt{\frac{pq}{n}}\,.\,t = \varepsilon$$

$$\therefore \quad P\left\{\left|\frac{m}{n} - p\right| \geq \varepsilon\right\} \leq \frac{pq}{n\varepsilon^2}$$

or $$P\left\{\left|\frac{m}{n} - p\right| \geq \varepsilon\right\} < \delta$$

When $$n > \frac{pq}{\varepsilon^2 \delta}$$

Then N = integer $> \dfrac{pq}{\varepsilon^2 \delta}$

EXERCISES

1. Find the variance of the distribution.

$$dF = \frac{1}{\pi} x \sin x \qquad 0 \leq x \leq \pi \qquad \left[\textbf{Ans. } 2 - \frac{16}{\pi^2}\right]$$

2. If $f(x) = be^{-bx}$, $0 < x < \infty$ where b is a positive constant. Find mean, μ_2 and μ_3.

$$\left[\textbf{Ans. } \frac{1}{b}, \frac{1}{b^2}, \frac{2}{b^3}\right]$$

3. Find μ_2, μ_3 and μ_4 for the distribution

$$dF = \frac{dx}{2a} \qquad -a \leq x \leq a \qquad \left[\textbf{Ans. } \frac{a^2}{3}, 0, \frac{a^4}{5}\right]$$

4. For the distribution

$$dF = x^m \frac{e^{-x}}{m!} \qquad m \geq 0,\ 0 \leq x < \infty$$

show that H.M. is m.

5. Find mean, mode and median of the distribution

$$dF = \sin x\, dx \qquad 0 \leq x \leq \frac{\pi}{2} \qquad \left[\textbf{Ans. } 1; \frac{\pi}{2}; \frac{\pi}{3}\right]$$

6. Find the mode and the medi

$y =$

7. Find the moment generating

$f(x) =$

and deduce the mean and va

8. Show that if x is a random v

$P\{a \leq x \leq b\} =$

then $a \leq E(x) \leq b$ and var (x

9. Find the mean deviation fro

$dF =$

and show that it is minimun

10. x is a random variable with

'a' is the median.

11. Does there exist a random v

P

Sol. By Tchebycheff's inec

P

Put $r =$

$\Rightarrow \quad P\{\bar{x} - 2\sigma_x \leq x \leq \bar{x} +$

i.e., $0{\cdot}6 \geq 0{\cdot}75$ which i

$\therefore$ there does not exist

12. **If X** is a random variate wit

and $P(X \leq 0) = 0$, show

$P(X > 2\mu) \leq \dfrac{1}{2}$

13. For the distribution given b

$f(x) =$

find the mean and variance

14. Let x be a random variabl

minimized when $b = \mu$.

15. For the continuous distribu

$f(x) =$

find $E(x)$ and var(x).

16. For a random variable wit

$F(x) =$

$\left[\text{Ans. } 2 - \dfrac{16}{\pi^2}\right]$

Find mean, μ_2 and μ_3.

$\left[\text{Ans. } \dfrac{1}{b}, \dfrac{1}{b^2}, \dfrac{2}{b^3}\right]$

$\left[\text{Ans. } \dfrac{a^2}{3}, 0, \dfrac{a^4}{5}\right]$

$: x \le \dfrac{\pi}{2}$ $\left[\text{Ans. } 1; \dfrac{\pi}{2}; \dfrac{\pi}{3}\right]$

6. Find the mode and the median of the curve

$$y = \frac{abx^{a-1}}{(1+bx^a)^2} \quad b>0,\ a>1,\ 0 \le x < \infty$$

$$\left[\text{Ans. } \left\{\frac{a-1}{b(a+1)}\right\}^{1/a};\ \left(\frac{1}{b}\right)^{1/a}\right]$$

7. Find the moment generating f^n of the distribution

$$f(x) = \frac{1}{2} e^{-|x|}, -\infty < x < \infty$$

and deduce the mean and variance.

8. Show that if x is a random variable such that

$$P\{a \le x \le b\} = 1$$

then $a \le E(x) \le b$ and var $(x) \le (b-a)^2$.

9. Find the mean deviation from 'a' of the distribution

$$dF = e^{-x}\,dx, \qquad x > 0$$

and show that it is minimum when 'a' equals the median of the distribution.

10. x is a random variable with a probability density. Show that $E|x-a|$ is minimum when 'a' is the median.

11. Does there exist *a* random variable x for which

$$P\{\bar{x} - 2\sigma_x \le x \le \bar{x} + 2\sigma_x\} = 0{\cdot}6$$

Sol. By Tchebycheff's inequality

$$P\{|x - \bar{x}| \le r\sigma_x\} \ge 1 - \frac{1}{r^2}$$

Put $\qquad r = 2$

$$\Rightarrow \quad P\{\bar{x} - 2\sigma_x \le x \le \bar{x} + 2\sigma_x\} \ge 1 - \frac{1}{4} = \frac{3}{4} = 0{\cdot}75$$

i.e., $\quad 0{\cdot}6 \ge 0{\cdot}75$ which is not possible.

$\therefore$ $\quad$ there does not exist any random variable with given prob.

12. If X is a random variate with $E(X) = \mu$
and $\quad P(X \le 0) = 0$, show that

$$P(X > 2\mu) \le \frac{1}{2}$$

13. For the distribution given by

$$f(x) = |1-x|\, I_{(0,\,2)}(x)$$

find the mean and variance. $\qquad \left[\text{Ans. mean} = 1, \text{ variance} = \dfrac{1}{2}\right]$

14. Let x be a random variable with mean μ and variance σ^2. Show that $E(x-b)^2$ is minimized when $b = \mu$.

15. For the continuous distribution with *p.d.f.*

$$f(x) = \lambda e^{-\lambda x} I_{(0,\,\infty)}$$

find $E(x)$ and var(x). $\qquad \left[\text{Ans. } \left(\dfrac{1}{\lambda}, \dfrac{1}{\lambda^2}\right)\right]$

16. For a random variable with *c.d.f*

$$F(x) = (1 - pe^{-\lambda x})\, I_{(0,\,\infty)}(x)$$

find mean and variance. $\left[\textbf{Ans. } \frac{p}{\lambda}, \frac{p(2-p)}{\lambda^2}\right]$

17. If x is a random variate for which

$$P(x \le 0) = 0 \qquad \text{and } E(x) = \mu < \infty,$$

show that

$$P[x \le \mu t] \ge 1 - \frac{1}{t}, \text{ for every } t \ge 1$$

Sol. Assume x is a continuous variate with *p.d.f.* $f(.)$.

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

$$= \int_{x \ge \mu t} x f(x) dx + \int_{x \le \mu t} x f(x) dx$$

$$\ge \int_{x \ge \mu t} x f(x) dx$$

$$\ge \int_{x \ge \mu t} \mu t f(x) dx = \mu t P[x > \mu t]$$

$$\Rightarrow \quad P(x > \mu t) \le \frac{1}{t}$$

$$\Rightarrow \quad 1 - P(x \le \mu t) \le \frac{1}{t}$$

$$P(x \le \mu t) \ge 1 - \frac{1}{t}.$$

□□

Theor

10.1. Binomial Distribution (B

Binomial Probability Dist

$P(x)$

The variate x is called **Bin**
the distribution.

Binomial Frequency Distr

$F(x)$

where N is the total frequency.

Derivation

Let there be N sets of n ind
trial be p and of failure is q. The

Let us first calculate the cha
Let us find the probability of ob
theorem of compound probabili

and the probability that the rem

∴ The probability of joint
trials failures

Clearly this is also the prob
any particular definite specified
and x trials can be chosen out of
probability, the probability of x

The chance of getting x succ

one set $^n c_x p^x q^{n-x}$ sets will hav

$$\left[\textbf{Ans. } \frac{p}{\lambda}, \frac{p(2-p)}{\lambda^2}\right]$$

$\iota < \infty$,

$(x)\,dx$

$> \mu t]$

□□

10

Theoretical Distribution

10.1. Binomial Distribution (B.D.)

Binomial Probability Distribution. *The B.P.D. of the variate x is*

$$P(x) = {}^n c_x p^x q^{n-x},\ x = 0, 1, 2, \ldots\ldots\ldots n$$

The variate x is called **Binomial Variate** (*B.V.*) *and n and p are called parameters of the distribution.*

Binomial Frequency Distribution. *The B.F.D. of the variate x is*

$$F(x) = N\, {}^n c_x p^x q^{n-x},\ x = 0, 1, 2, \ldots\ldots n$$

where N is the total frequency.

Derivation

Let there be N sets of n independent trials. Assume that the chance of success of each trial be p and of failure is q. Then

$$p + q = 1.$$

Let us first calculate the chances of obtaining 0, 1, 2,.....successes in one set of n trials. Let us find the probability of obtaining x successes and $(n-x)$ failures in n trials. By the theorem of compound probability, the probability that first x trials are successes

$$= p \times p \times p..x \text{ times} = p^x$$

and the probability that the remaining $(n-x)$ trials are failures

$$= q^{n-x}$$

$\therefore$ The probability of jointly getting first x trials success and the remaining $(n-x)$ trials failures

$$= p^x q^{n-x}$$

Clearly this is also the probability for the x successes and $(n-x)$ failures to occur in any particular definite specified order. Since we are interested in any x trials being successes and x trials can be chosen out of n in ${}^n c_x$ (mutually exclusive) ways, by the theorem of total probability, the probability of x successes in a series of n trials is given by

$$P(x) = {}^n c_x p^x q^{n-x}$$

The chance of getting x successes in one set of n trials is ${}^n c_x p^x q^{n-x}$ means that out of one set ${}^n c_x p^x q^{n-x}$ sets will have x successes.

$\therefore$ Out of N sets $N\ {}^nc_x p^x q^{n-x}$ sets will have x successes.

$\therefore$ The frequencies of getting 0, 1, 2,...successes in N sets of n trials each are $Nq^n, N\,{}^nc_1 pq^{n-1}, N\ {}^nc_2 p^2 q^{n-2} \ldots\ldots N\ {}^nc_n p^n$.

10.1.1. First Four Moments About Mean

Binomial Probability distribution is

$$P(x) = {}^nc_x p^x q^{n-x},\ x = 0, 1, \ldots n$$

$$\begin{aligned}
\mu_1'(0) &= \sum_{x=0}^{n} xP(x) = \sum_{x=0}^{n} x\ {}^nc_x p^x q^{n-x} \\
&= {}^nc_1 pq^{n-1} + 2\ {}^nc_2\ p^2 q^{n-2} + \ldots + n\ {}^nc_n p^n \\
&= np\,\{q^{n-1} + (n-1)\,pq^{n-2} + \ldots + p^{n-1}\} \\
&= np\,(q+p)^{n-1} \\
&= np
\end{aligned}$$

$$\begin{aligned}
\mu_2'(0) &= \sum_{x=0}^{n} x^2 P(x) = \sum_{x=0}^{n} \{x(x-1) + x\}\ {}^nc_x p^x q^{n-x} \\
&= \sum_{x=0}^{n} x(x-1)\ {}^nc_x p^x q^{n-x} + \sum_{x=0}^{n} x\ {}^nc_x p^x q^{n-x} \\
&= 2.1\ {}^nc_2 p^2 q^{n-2} + 3.2\ {}^nc_3 p^3 q^{n-3} + \ldots + n(n-1)p^n + np \\
&= n(n-1)\,p^2\,(q^{n-2} + (n-2)\,pq^{n-3} + \ldots + p^{n-2}) + np \\
&= n(n-1)\,p^2\ (q+p)^{n-2} + np \\
&= n(n-1)\,p^2 + np
\end{aligned}$$

$$\begin{aligned}
\mu_2 &= \mu_2'(0) - \{\mu_1'(0)\}^2 = n(n-1)p^2 + np - n^2p^2 \\
&= np\,(1-p) = npq
\end{aligned}$$

$$\mu_3'(0) = \sum_{x=0}^{n} x^3 P(x)$$

Put $\qquad x^3 \equiv Ax + Bx(x-1) + Cx(x-1)\,(x-2)$

Equating co-efficients of x^3, x^2 and x

$$C = 1$$

$$B - 3C = 0 \ \Rightarrow\ B = 3$$

and $\qquad A - B + 2C = 0 \Rightarrow A = 1$

$\therefore \qquad x^3 \equiv x + 3x(x-1) + x(x-1)(x-2)$

μ_3'

$\therefore$

μ_4'

$x^4 \equiv x(x-1)(x-2)(x-3)$

$\therefore\ \mu_4'(0) = \sum_{x=0}^{n} \{x(x-1)(x$

$= n(n-1)(n-2)(n-3)p^4 +$

$\therefore\ \mu_4 = \mu_4'(0) - 4\mu_3'(0)\mu_1'($

$= \{n(n-1)(n-2)(n-3)p^4$

$-4\{n(n-1)(n-2)p^3 + 3n(n$

$= (n^4 - 6n^3 + 11n^2 - 6n)p^4$

$-(n^4 - 3n^3 + 2n^2)p^4 - 12($

$= 3n^2p^4 - 6np^4 - 6n^2p^3 +$

$= np\{(1 - 7p + 12p^2 - 6p^3)$

$= np\{(1-p)(1-6p+6p^2)$

$= npq\ \{1 - 6p(1-p) + 3npq$

$= npq\ \{1 + 3pq(n-2)\}$

Ex. 10-1. *For binomial dist*

$$\mu_3'(0) = \sum_{x=0}^{n} \{x + 3x(x-1) + x(x-1)(x-2)\}\ {}^n n_x p^x q^{n-x}$$

$$= np + 3n(n-1)p^2 + n(n-1)(n-2)p$$

$$\therefore \qquad \mu_3 = \mu_3'(0) - 3\mu_2'(0)\,\mu_1'(0) + 2\,\{\mu_1'(0)\}^3$$

$$= np + 3n(n-1)p^2 + n(n-1)(n-2)p^3$$

$$-3\{n(n-1)p^2 + np\}np + 2n^3p^3$$

$$= np + 3n^2p^2 - 3np^2 + n^3p^3 - 3n^2p^3 + 2np^3$$

$$-3n^3p^3 + 3n^2p^3 - 3n^2p^2 + 2n^3p^3$$

$$= np\,(1 - 3p + 2p^2)$$

$$= np(1-p)(1-2p)$$

$$= npq(q-p)$$

$$\mu_4'(0) = \sum_{x=0}^{n} x^4 P(x)$$

$$x^4 \equiv x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x$$

$$\therefore\ \mu_4'(0) = \sum_{x=0}^{n} \{x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x\}.{}^n c_x p^x q^{n-x}$$

$$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

$$\therefore\ \mu_4 = \mu_4'(0) - 4\mu_3'(0)\mu_1'(0) + 6\mu_2'(0)\{\mu_1'(0)\}^2 - 3\{\mu_1'(0)\}^4$$

$$= \{n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np\}$$

$$-4\{n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np\}np + 6\{n(n-1)p^2 + np\}n^2p^2 - 3n^4p^4$$

$$= (n^4 - 6n^3 + 11n^2 - 6n)p^4 + 6(n^3 - 3n^2 + 2n)p^3 + 7(n^2 - n)p^2 + np$$

$$-(n^4 - 3n^3 + 2n^2)p^4 - 12(n^3 - n^2)p^3 - 4n^2p^2 + 6(n^4 - n^3)p^4 + 6n^3p^3 - 3n^4p^4$$

$$= 3n^2p^4 - 6np^4 - 6n^2p^3 + 12np^3 + 3n^2p^2 - 7np^2 + np$$

$$= np\{(1 - 7p + 12p^2 - 6p^3) + 3np(1 - 2p + p^2)\}$$

$$= np\{(1-p)(1 - 6p + 6p^2) + 3np(1-p)^2\}$$

$$= npq\,\{1 - 6p(1-p) + 3npq\}$$

$$= npq\,\{1 + 3pq(n-2)\}$$

Ex. 10-1. *For binomial distribution show that*

$$\mu_{r+1} = pq\left\{n.r.\mu_{r-1} + \frac{d\mu_r}{dp}\right\}$$

and deduce the values of μ_2, μ_3 *and* μ_4.

Sol. Binomial distribution is

$$P(x) = {}^n c_x p^x q^{n-x}, \; x = 0, 1, 2, \ldots n$$

By d.f.

$$\mu_r = E(x-np)^r$$

$$= \sum_{x=0}^{n} {}^n c_x p^x q^{n-x} (x-np)^r$$

$$\therefore \quad \frac{d\mu_r}{dp} = \sum_{x=0}^{n} {}^n c_x \{x p^{x-1} q^{n-x} (x-np)^r$$

$$+ p^x (n-x) q^{n-x-1} \frac{dq}{dp} (x-np)^r + p^x q^{n-x} r (x-np)^{r-1} (-n)\}$$

$$= \sum_{x=0}^{n} {}^n c_x p^{x-1} q^{n-x-1} (x-np)^r \{xq - p(n-x)\}$$

$$-nr \sum_{x=0}^{n} {}^n c_x p^x q^{n-x} (x-np)^{r-1} \qquad \left(\because \frac{dq}{dp} = -1\right)$$

$$= \frac{1}{pq} \sum_{x=0}^{n} {}^n c_x p^x q^{n-x} (x-np)^{r+1} - nr\,\mu_{r-1}$$

$$= \frac{1}{pq} \mu_{r+1} - nr\mu_{r-1}$$

$$\therefore \quad \mu_{r+1} = pq\left\{nr\mu_{r-1} + \frac{du_r}{dp}\right\}$$

Put

$$r = 1, 2 \text{ and } 3$$

$$\mu_2 = pq\left\{n\mu_0 + \frac{d\mu_1}{dp}\right\} = npq \qquad (\because \mu_0 = 1, \mu_1 = 0)$$

$$\mu_3 = pq\left\{2n\mu_1 + \frac{d\mu_2}{dp}\right\} = npq\frac{d}{dp}\{pq\}$$

$$= npq(q-p)$$

$$\mu_4 = pq\left\{3n\mu_2 + \frac{d\mu_3}{dp}\right\} = npq\left[3npq + \frac{d}{dp}\{pq(q-p)\}\right]$$

$$= npq[3npq + (q-p)^2 - 2pq]$$

$$= npq[(q+p)^2 + 3pq(n-2)]$$

$$= npq\,[1 + 3pq(n-2)].$$

Ex. 10-2. *For a binomial varia*

μ'_{r+1}

where $\mu'_r = E(x^r)$ *and r is a non-n*

Sol. By def.

μ'_r

$\therefore \quad \frac{d\mu'_r}{dp}$

μ'_{r+1}

10.1.2. Measures of Skewness

Sol. Measures of skewness

Measure of kurtosis

$2, \ldots n$

$p)^r$

$-np)^r$

$n)\}$

$\left(\because \frac{dq}{dp} = -1\right)$

$pq \qquad (\because \mu_0 = 1, \mu_1 = 0)$

$npq \frac{d}{dp} \{pq\}$

$npq(q-p)$

$npq\left[3npq + \frac{d}{dp}\{pq(q-p)\}\right]$

$-2pq]$

$n-2)]$

Ex. 10-2. *For a binomial variate x with parameters n and p show that*

$$\mu'_{r+1} = np\mu'_r + pq\frac{d\mu'_r}{dp}$$

where $\mu'_r = E(x^r)$ *and r is a non-negative integer.*

Sol. By def.

$$\mu'_r = E(x^r)$$

$$= \sum_{x=0}^{n} x^r \, {}^nc_x p^x q^{n-x}$$

$$\therefore \quad \frac{d\mu'_r}{dp} = \sum_{x=0}^{n} x^r \, {}^nc_x [xp^{x-1}q^{n-x} - (n-x)q^{n-x-1}p^x]$$

$$\left(\because \frac{dq}{dp} = -1\right)$$

$$= \sum_{x=0}^{n} x^r \, {}^nc_x p^{x-1} q^{n-x-1} [xq - (n-x)p]$$

$$= \frac{1}{pq} \sum_{x=0}^{n} x^r \, {}^nc_x \, p^x q^{n-x} \, (x - np)$$

$$= \frac{1}{pq} \left\{ \sum_{x=0}^{n} x^{r+1} \, {}^nc_x p^x q^{n-x} - np \sum_{x=0}^{n} x^r \, {}^nc_x p^x q^{n-x} \right\}$$

$$= \frac{1}{pq} \{\mu'_{r+1} - np\mu'_r\}$$

$$\mu'_{r+1} = np\,\mu'_r + pq\frac{d\mu'_r}{dp}.$$

10.1.2. Measures of Skewness and Kurtosis

Sol. Measures of skewness

$$= \gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}}$$

$$= \frac{\mu_3}{\mu_2^{3/2}}$$

$$= \frac{npq\,(q-p)}{(npq)^{3/2}} = \frac{q-p}{\sqrt{npq}}$$

Measure of kurtosis $= \beta_2 = \dfrac{\mu_4}{\mu_2^2}$

$$= \frac{npq\{1+3pq(n-2)\}}{(npq)^2}$$

$$= \frac{1+3pq(n-2)}{npq} = 3+\frac{1-6pq}{npq}$$

$$\therefore \qquad \gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}.$$

10.1.3. Mean Deviation about Mean

Mean deviation about mean is given by

$$\eta = E|x-np|$$

$$= \sum_{x=0}^{n} |x-np|\ {}^nc_x p^x q^{n-x}$$

$$= \sum_{x>np} (x-np)\ {}^nc_x p^x q^{n-x} + \sum_{x<np} (np-x)\ {}^nc_x p^x q^{n-x}$$

Now $\mu_1 = 0$

$$\therefore \sum_{x=0}^{n} (x-np)\ {}^nc_x\ p^x q^{n-x} = 0$$

$$\Rightarrow \sum_{x>np} (x-np)\ {}^nc_x p^x q^{n-x} = \sum_{x<np} (np-x)\ {}^nc_x\ p^x q^{n-x}$$

$$\therefore \qquad \eta = 2\sum_{x>np} (x-np)\ {}^nc_x\ p^x q^{n-x}$$

$$= 2\sum_{x>np} (xq-(n-x)p)\ {}^nc_x\ p^x q^{n-x}$$

$$= 2\sum_{x>np} \{x\ {}^nc_x p^x q^{n-x+1} - (n-x)\ {}^nc_x p^{x+1} q^{n-x}\}$$

$$= 2\sum_{x>np} \left\{\frac{n!}{(x-1)!\,(n-x)!} p^x\ q^{n-x+1} - \frac{n!}{x!\,(n-x-1)!}\ p^{x+1} q^{n-x}\right\}$$

$$= 2\sum_{x>np} \{t_{x-1} - t_x\}$$

where $\qquad t_x = \dfrac{n!}{x!(n-x-1)!}\ p^{x+1} q^{n-x}.$

Let $\mu =$ greatest integer contained in $np+1$

Then $\qquad \eta = 2\sum_{x=\mu}^{n} \{t_{x-1} - t_x\}$

10.1.4. Mode of the Binomi

In binomial distribution the

$P(x)$

The values of x together with

The mode is that value of x for wh

i.e.,

$P(x-1) \le P(x$

Consider $\quad P(x-1) \le P(x)$

or $\quad {}^nc_{x-1}\ p^{x-1} q^{n-}$

or $\quad \dfrac{n!}{(x-1)!\,(n-x+}$

or $\quad xq \le (n+1)\ p -$

or $\quad x(q+p) \le (n+1)$

or $\quad x \le (n+1)p$

Similarly other inequality giv

$x \ge (n+1)p-1$

From (*i*) and (*ii*), modal value

$(n+1)p-1 \le x \le$

Case I : If $(n+1)\,p = k$ is an

Now $\quad \dfrac{P(x=k)}{P(x=k-1}$

$\therefore \quad P(x=k$

$\therefore$ In this case $P(x)$ increases t decrease.

$\therefore x = k$ and $x = k-1$ are two

Case II : If $(n+1)\ p \neq k$ is no

$(n+1)\ p = a$

$$= 2\{t_{\mu-1} - t_n\}$$

$$= 2t_{\mu-1} \qquad (\because t_n = 0)$$

$$= \frac{2.n!}{(\mu-1)!(n-\mu)!} p^{\mu} q^{n-\mu+1}$$

$$= 2npq\,\{{}^{n-1}c_{\mu-1} p^{\mu-1} q^{n-\mu}\}.$$

10.1.4. Mode of the Binomial Distribution

In binomial distribution the probability of x successes is given by

$$P(x) = {}^n c_x\, p^x q^{n-x},\ x = 0, 1, 2, \ldots n$$

The values of x together with the corresponding probabilities form binomial distribution. The mode is that value of x for which $P(x)$ is greater than or equal to $P(x-1)$ and $P(x+1)$ *i.e.*,

$$P(x-1) \le P(x) \ge P(x+1)$$

Consider $\quad P(x-1) \le P(x)$

or $\quad {}^n c_{x-1}\, p^{x-1} q^{n-x+1} \le {}^n c_x\, p^x q^{n-x}$

or $\quad \dfrac{n!}{(x-1)!\,(n-x+1)!}\, q \le \dfrac{n!}{x!\,(n-x)!}\, p$

or $\quad xq \le (n+1)\,p - xp$

or $\quad x(q+p) \le (n+1)p$

or $\quad x \le (n+1)p \qquad \ldots(i)$

Similarly other inequality gives

$$x \ge (n+1)p - 1 \qquad \ldots(ii)$$

From (*i*) and (*ii*), modal value x satisfies the inequality

$$(n+1)p - 1 \le x \le (n+1)p \qquad \ldots(iii)$$

Case I : If $(n+1)\,p = k$ is an integer, then $(n+1)p - 1 = k - 1$ is also an integer.

Now $\quad \dfrac{P(x=k)}{P(x=k-1)} = \dfrac{{}^n c_k\, p^k q^{n-k}}{{}^n c_{k-1}\, p^{k-1} q^{n-k+1}}$

$$= \frac{n!}{k!\,(n-k)!} \cdot \frac{(k-1)!\,(n-k+1)!}{n!}\, \frac{p}{q}$$

$$= \frac{(n+1)\,p - kp}{kq} = \frac{k(1-p)}{kq} = 1$$

$\therefore \quad P(x=k) = P(x=k-1) \qquad \ldots(iv)$

$\therefore$ In this case $P(x)$ increases till $x = k-1$ and then (*iv*) holds and after that it begins to decrease.

$\therefore x = k$ and $x = k-1$ are two modes.

Case II : If $(n+1)\,p \neq k$ is not an integer, let

$$(n+1)\,p = a \text{ (an integer)} + f (a \text{ fraction})$$

when x takes the vale 'a' (which is obviously less than k and greater than $k-1$) from (i) and (ii)

$$P(a-1) < P(a) > P(a+1)$$

$\therefore$ $x = a$ (greatest integer less than k) is the mode.

Ex. 10-3. *If np be a whole number, the mean of the binomial distribution coincides with the greatest term.*

Sol. If np is a whole number, then since p is a fraction, np is the greatest integer less than $np+p=k$.

$\therefore$ From Case II, mode $= np =$ mean.

Ex. 10-3. (a), *If x is the unique mode of the B.D., show that*

$$(n+1)p-1 < x < (n+1)p.$$

10.1.5. Moment Generating Function

M.G.F., by def. is given by

$$M_0(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx}\ {}^nc_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nc_x\,(pe^t)^x q^{n-x} = (q+pe^t)^n$$

M.G.F. about the mean 'np' is given by

$$M_{\bar{x}}(t) = E(e^{t(x-np)}) = e^{-npt} M_0(t)$$

$$= e^{-npt}(q+pe^t)^n$$

$$= \{qe^{-pt} + pe^{qt}\}^n.$$

Deduction of moments about mean

$$M_{\bar{x}}(t) = (qe^{-pt} + pe^{qt})^n$$

$$= \left\{q\left(1-pt+p^2\frac{t^2}{2!}-p^3\frac{t^3}{3!}+p^4\frac{t^4}{4!}+\ldots\right)+p\left(1+qt+q^2\frac{t^2}{2!}+q^3\frac{t^3}{3!}+q^4\frac{t^4}{4!}+\ldots\right)\right\}^n$$

$$= \left\{1+pq\frac{t^2}{2!}+pq(q^2-p^2)\frac{t^3}{3!}+qp(p^3+q^3)\frac{t^4}{4!}+\ldots\right\}^n$$

$$= \left\{1+pq\frac{t^2}{2!}+pq(q-p)\frac{t^3}{3!}+qp(p^2-pq+q^2)\frac{t^4}{4!}+\ldots\right\}^n$$

$$\therefore\ 1+\mu_1 t+\mu_2\frac{t^2}{2!}+\mu_3\frac{t^3}{3!}+\mu_4\frac{t^4}{4!}+\ldots$$

$$=1+npq\frac{t^2}{2!}+npq\,(q-p)\frac{t^3}{3!}+[npq\,\{p^2-pq+q^2\}+3n(n-1)p^2q^2]\frac{t^4}{4!}+\ldots$$

$\therefore$ $\mu_1 = 0, \mu_2 = npq, \mu_3 = npq\,(q-p)$

and

Ex. 10-4. *Is the sum of two*
what are the conditions under w

Sol. Let x_1 and x_2 be two
n_2, p_2 respectively.

Then $M_0(t)$ of

and $M_0(t)$ of

Let

Then $M_0(t)$ o

which being not of the form $(q$ -

If $p_1 = p_2 = p$ so that $q_1 =$

Then $M_0(t)$ of $x = (q + pe$

which implies that x is a binomi
required condition is

10.1.6. Cumulative Functi

By def. cumulative f^n is g

$K_0(t) = \log$

$= n \log$

$= n \log$

$= n$

But $K_0(t) = k_1 t$ -

$\therefore$ $k_1(0) = np, k$

and $$\mu_4 = npq\{p^2 - pq + q^2 + 3(n-1)pq\}$$
$$= npq\{1 + 3(n-2)pq\}.$$

Ex. 10-4. *Is the sum of two independent binomial variates a binomial variate ? If not, what are the conditions under which it is so ?*

Sol. Let x_1 and x_2 be two independent binomial variates with parameters n_1, p_1 and n_2, p_2 respectively.

Then $$M_0(t) \text{ of } x_1 = (q_1 + p_1 e^t)^{n_1}$$

and $$M_0(t) \text{ of } x_2 = (q_2 + p_2 e^t)^{n_2}$$

Let $$x = x_1 + x_2$$

Then $$M_0(t) \text{ of } x = \{M_0(t) \text{ of } x_1\}.\{M_0(t) \text{ of } x_2\}$$
$$= (q_1 + p_1 e^t)^{n_1}.(q_2 + p_2 e^t)^{n_2}$$

which being not of the form $(q + pe^t)^n$ implies that x is not a binomial variate.

If $p_1 = p_2 = p$ so that $q_1 = q_2 = q$

Then $M_0(t)$ of $x = (q + pe^t)^{n_1 + n_2}$

which implies that x is a binomial variate with parameters $(n_1 + n_2)$ and p. Therefore, the required condition is
$$p_1 = p_2.$$

10.1.6. Cumulative Function and Cumulants

By def. cumulative f^n is given by

$$K_0(t) = \log M_0(t) = n \log(q + pe^t)$$
$$= n \log\left\{q + p\left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \ldots\right)\right\}$$
$$= n \log\left\{1 + pt + p\frac{t^2}{2!} + p\frac{t^3}{3!} + p\frac{t^4}{4!} + \ldots\right\}$$
$$= n\left[\left\{pt + p\frac{t^2}{2!} + p\frac{t^3}{3!} + p\frac{t^4}{4!} + \ldots\right\} - \frac{1}{2}\left\{pt + p\frac{t^2}{2!} + p\frac{t^3}{3!} + \ldots\right\}^2\right.$$
$$\left. + \frac{1}{3}\left\{pt + p\frac{t^2}{2!} + \ldots\right\}^3 - \frac{1}{4}\{pt + \ldots\}^4 + \ldots\ldots\right]$$

But $$K_0(t) = k_1 t + k_2\frac{t^2}{2!} + k_3\frac{t^3}{3!} + \ldots\ldots$$

$\therefore$ $$k_1(0) = np, k_2 = npq,$$

$$k_3 = n\{p-3p^2+2p^3\} = np\,\{1-3p+2p^2\}$$
$$= np(1-2p)(1-p) = npq\,(q-p)$$
$$k_4 = n[p-7p^2+12p^3-6p^4] = np\,\{1-7p+12p^2-6p^3]$$
$$= np\{(1-p)(1-6p+6p^2)\}$$
$$= npq\{1-6p(1-p)\}$$
$$= npq\,\{1-6pq\}.$$

Ex. 10-5. *Show that for the binomial dist. with parameters n and p.*

$$k_{r+1} = pq\frac{dk_r}{dp}$$

Hence deduce the values of k_2, k_3 and k_4.

Sol. For B.D., $M_0(t) = (q+pe^t)^n$

$$\therefore \quad K_0(t) = \log M_0(t) = n.\log\{q+pe^t\}$$

$$\therefore \quad k_r = n\left[\frac{d^r}{dt^r}\{\log(q+pe^t)\}\right]_{t=0}$$

$$\therefore \quad \frac{dk_r}{dp} = n\left[\frac{d^r}{dt^r}\left\{\frac{e^t-1}{q+pe^t}\right\}\right]_{t=0}$$

Also
$$k_{r+1} = n\left[\frac{d^{r+1}}{dt^{r+1}}\{\log(q+pe^t)\}\right]_{t=0}$$

$$= n\left[\frac{d^r}{dt^r}\left\{\frac{pe^t}{q+pe^t}\right\}\right]_{t=0}$$

$$\therefore \quad k_{r+1} - pq\frac{dk_r}{dp} = n\left[\frac{d^r}{dt^r}\left\{\frac{pe^t-pqe^t+pq}{q+pe^t}\right\}\right]_{t=0}$$

$$= \left[\frac{d^r}{dt^r}\left\{\frac{p(pe^t+q)}{pe^t+q}\right\}\right]_{t=0}$$

$$= n\left[\frac{d^r}{dt^r}(p)\right]_{t=0} = 0$$

$$\therefore \quad k_{k+1} = pq\frac{dk_r}{dp}$$

Put $r = 1, 2,$ and 3

and

Ex. 10-6. *If a coin is tossed*

probability of exactly $\frac{n}{2} - x$ head.

Sol. Let the occurrence of a h

Then

$\therefore$ Probability of x successes

$P($.

Since n is an even number, le

where k is a positive integer.

$\therefore$ $P($.

Now $\frac{P(k-x}{p(k)}$

By Stirling's formula.

$$k_2 = pq\frac{dk_1}{dp} = npq \text{ as } k_1 = \mu_1'(0) = np$$

$$k_3 = pq\frac{dk_2}{dp} = npq\,(q-p)$$

and
$$k_4 = pq\frac{dk_3}{dp} = npq\,\{q(q-p)-p(q-p)-2pq\}$$
$$= npq\,\{(q+p)^2-6pq\}$$
$$= npq\,\{1-6pq\}$$

Ex. 10-6. *If a coin is tossed n times where n is a large even number, show that the probability of exactly* $\frac{n}{2}-x$ *heads and* $\frac{n}{2}+x$ *tails is*

$$\left(\frac{2}{\pi n}\right)^{\frac{1}{2}}.e^{-\frac{2x^2}{n}}$$

Sol. Let the occurrence of a head in a toss be called success and p be its probability.

Then
$$p = \frac{1}{2} = q$$

$\therefore$ Probability of x successes is given by

$$P(x) = {}^{n}c_x.\left(\frac{1}{2}\right)^n$$

Since n is an even number, let
$$n = 2k$$
where k is a positive integer.

$$\therefore \qquad P(x) = {}^{2k}c_x\left(\frac{1}{2}\right)^{2k}$$

Now
$$\frac{P(k-x)}{p(k)} = \frac{{}^{2k}c_{k-x}\left(\frac{1}{2}\right)^{2k}}{{}^{2k}c_k\left(\frac{1}{2}\right)^{2k}}$$

$$= \frac{(2k)!}{(k-x)!\,(k+x)!}\,\frac{k!\,k!}{(2k)!}$$

$$= \frac{k!\,k!}{(k-x)!\,(k+x)!}$$

By Stirling's formula.

$$k! \simeq \sqrt{2\pi}\;e^{-k}\;k^{k+\frac{1}{2}}$$

$$\therefore \quad \frac{P(k-x)}{P(k)} = \frac{\sqrt{2\pi}e^{-k}\,k.^{k+\frac{1}{2}}.\sqrt{2\pi}e^{-k}.k^{k+\frac{1}{2}}}{\sqrt{2\pi}e^{-k+x}(k-x)^{k-x+\frac{1}{2}}.\sqrt{2\pi}e^{-k-x}(k+x)^{k+x+\frac{1}{2}}}$$

$$= \frac{k^{2k+1}}{(k-x)^{k-x+\frac{1}{2}}\,(k+x)^{k+x+\frac{1}{2}}}$$

$$= \frac{1}{\left(1-\frac{x}{k}\right)^{k-x+\frac{1}{2}}\left(1+\frac{x}{k}\right)^{k+x+\frac{1}{2}}}$$

$$\therefore \quad \log\frac{P(k-x)}{P(k)} \approx -\left(k-x+\frac{1}{2}\right)\log\left(1-\frac{x}{k}\right) - \left(k+x+\frac{1}{2}\right)\log\left(1+\frac{x}{k}\right)$$

$$= \left(k-x+\frac{1}{2}\right)\left(\frac{x}{k}+\frac{1}{2}\frac{x^2}{k^2}+\ldots\ldots\right) - \left(k+x+\frac{1}{2}\right)\left(\frac{x}{k}-\frac{1}{2}\frac{x^2}{k^2}+\ldots\ldots\right)$$

$$= -\frac{x^2}{k}$$

neglecting terms containing $\frac{1}{k^2}$ and higher powers of $\frac{1}{k}$ as k is large.

$$\therefore \quad P(k-x) \approx P(k)e^{-\frac{x^2}{k}}$$

$$= {}^{2k}c_k.\left(\frac{1}{2}\right)^{2k}.e^{-\frac{2x^2}{n}}$$

$$= \frac{(2k)!}{k!k!}\left(\frac{1}{2}\right)^{2k}.e^{-\frac{2x^2}{n}}$$

$$= \frac{\sqrt{2\pi}e^{-2k}.(2k)^{2k+\frac{1}{2}}}{\left(\sqrt{2\pi}e^{-k}.k^{k+\frac{1}{2}}\right)^2}.\left(\frac{1}{2}\right)^{2k}.e^{-\frac{2x^2}{n}}$$

$$= \frac{1}{\sqrt{\pi k}}e^{-\frac{2x^2}{n}}$$

$$= \sqrt{\frac{2}{\pi n}}.e^{-\frac{2x^2}{n}}.$$

Ex. 10-7. *Six dice are throw*
dice to show a 5 or 6 ?

Sol. Here $N = 729$, $n = 6$

Let the occurrence of 5 or 6

Now p = prob. of occurrenc

$\therefore$

$\therefore$ Prob. of x successes is g

By theorem of total probability,

$\therefore$ No. of times at least thr

Ex. 10-8. *A perfect cubic di*
of 5 or 6 is called a success. In

Sol. Here

$\therefore$

$\therefore$ Probability of x success

$\therefore$ $P(x$

$$\frac{{}^{k+\frac{1}{2}}.\sqrt{2\pi}e^{-k}.k^{k+\frac{1}{2}}}{{}^{x+\frac{1}{2}}.\sqrt{2\pi}e^{-k-x}(k+x)^{k+x+\frac{1}{2}}}$$

$$\overline{\quad)^{k+x+\frac{1}{2}}}$$

$$\overline{\frac{x}{k}\Big)^{k+x+\frac{1}{2}}}$$

$$-\frac{x}{k}\Big)-\left(k+x+\frac{1}{2}\right)\log\left(1+\frac{x}{k}\right)$$

$$\Big)-\left(k+x+\frac{1}{2}\right)\left(\frac{x}{k}-\frac{1}{2}\frac{x^2}{k^2}+\ldots\ldots\right)$$

as k is large.

$$-\frac{2x^2}{n}$$

Ex. 10-7. *Six dice are thrown 729 times. How many times do you expect at least three dice to show a 5 or 6 ?*

Sol. Here $N = 729$, $n = 6$

Let the occurrence of 5 or 6 be regarded as success and p be the probability of success.

Now p = prob. of occurrence of 5 or 6

$$= \frac{2}{6} = \frac{1}{3}$$

$$\therefore \quad q = 1 - p = \frac{2}{3}$$

$\therefore$ Prob. of x successes is given by

$$p(x) = {}^6c_x p^x q^{6-x} = \frac{{}^6c_x\, 2^{6-x}}{3^6}$$

By theorem of total probability, probability of at least three successes

$$= P(3) + P(4) + P(5) + P(6)$$

$$= \frac{1}{729}\{{}^6c_3.2^3 + {}^6c_4.2^2 + {}^6c_5.2 + {}^6c_6\}$$

$$= \frac{1}{729}\left\{\frac{6.5.4}{3.2.1}\cdot 8 + \frac{6.5.4.3}{4.3.2.1}\cdot 4 + 6.2 + 1\right\}$$

$$= \frac{1}{729}\{160 + 60 + 12 + 1\}$$

$$= \frac{233}{729}$$

$\therefore$ No. of times at least three successes occur

$$= \frac{233}{729}\,729 = 233.$$

Ex. 10-8. *A perfect cubic die is thrown a large number of times in sets of 8. The occurrence of 5 or 6 is called a success. In what proportion of the sets would you expect 3 successes ?*

Sol. Here $n = 8$

p = probability of success

= probability of occurrence of 5 or 6

$$= \frac{2}{6} = \frac{1}{3}$$

$$\therefore \quad q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

$\therefore$ Probability of x successes is given by

$$P(x) = {}^8c_x\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{8-x} = \frac{{}^8c_x.2^{8-x}}{3^8}$$

$$\therefore \quad P(x=3) = \frac{{}^8c_3.2^5}{3^8} = \frac{8.7.6}{3.2}\cdot\frac{32}{81\times 81}$$

$$= \frac{1792}{6561} = 0 \cdot 2731.$$

$\therefore$ Required proportion $= 0{\cdot}2731 = 27{\cdot}31\%$.

Ex. 10-9. *Assuming that half the population are consumers of rice so that the chance of an individual being a consumer is $\frac{1}{2}$ and assuming that 100 investigators each take 10 individuals to see whether they are consumers, how many investigators would you expect to report that three people or less were consumers ?*

Sol. Here $\quad N = 100, n = 10, \; p = \frac{1}{2}$

$$\therefore \quad q = \frac{1}{2}$$

$$\therefore \quad P(x) = {}^{10}c_x \left(\frac{1}{2}\right)^{10}$$

$\therefore$ Required number $= 100 \; \{P(3) + P(2) + P(1) + P(0)\}$

$$= \frac{100}{2^{10}} \{{}^{10}c_3 + {}^{10}c_2 + {}^{10}c_1 + {}^{10}c_0\}$$

$$= \frac{100}{1024} \left\{ \frac{10.9.8}{3.2.1} + \frac{10.9}{2.1} + 10 + 1 \right\}$$

$$= \frac{100}{1024} \{120 + 45 + 10 + 1\}$$

$$= \frac{17600}{1024} \approx 17.$$

Ex. 10-10. *An irregular six-faced die is thrown and the expectation that in 10 throws it will give five even numbers is twice the expectation that it will give four even numbers. How many times in 10,000 sets of 10 throws would you expect it to give no even number ?*

Sol. Let p be the probability of an even number and $q = 1 - p$.

Then the prob. of getting five even numbers

$$= {}^{10}c_5 . p^5 q^5$$

and the prob. of getting four even numbers

$$= {}^{10}c_4 . p^4 q^6$$

By given,

$${}^{10}c_5 . p^5 q^5 = 2. \, {}^{10}c_4 . p^4 q^6$$

or $\quad 3p = 5q = 5 - 5p$

$$\therefore \quad p = \frac{5}{8} \text{ and } q = \frac{3}{8}$$

$$\therefore \text{ Required number} \quad = 10{,}000 \left\{ {}^{10}c_0 \left(\frac{3}{8}\right)^{10} \right\}$$

Ex. 10-11. *In a precision bom*
strike the target. Two direct hits
bombs must be dropped to give a

Sol. Here

Let n be the required number
order to destroy it completely.

$\therefore \quad P($

where $\quad P($

$\therefore \quad 1 - P($

$\therefore \quad 0{\cdot}01 \geq P($

or

The value of n is the least po
Putting

which is not true.
Putting

which is true.

$\therefore$

Ex. 10-12. *Show that if two s*

n (and of the same number of o
coincides with the $(r+1)$th term
terms is a symmetrical binomial

Sol. Let N be the number of
distribution are

$$N.\left(\frac{1}{2}\right)^n, N\,{}^nc_1\left(\right.$$

$\therefore$ rth term of the first distrib

and $(r+1)$th term of the second

's of rice so that the chance of

0 investigators each take 10

stigators would you expect to

$+\ {}^{10}c_0\}$

$10+1\}$

$1\}$

pectation that in 10 throws it
give four even numbers. How
give no even number ?
$-p$.

$$= 10{,}000\left(\frac{3}{8}\right)^{10}$$

$$= 0\cdot549 \approx 1.$$

Ex. 10-11. *In a precision bombing attack there is a 50% chance that any one bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% chance or better of completely destroying the target ?*

Sol. Here $$p = \frac{1}{2} = q.$$

Let n be the required number of bombs. Out of n at least 2 bombs must hit the target in order to destroy it completely.

$$\therefore \quad P(2) + P(3) + \ldots + P(n) \geq 0\cdot99$$

where $$P(x) = {}^nc_x\left(\frac{1}{2}\right)^n,$$

$$\therefore \quad 1 - P(0) - P(1) \geq 0\cdot99$$

$$\therefore \quad 0\cdot01 \geq P(0) + P(1) = \frac{1}{2^n} + \frac{{}^nc_1}{2^n}$$

or $$100\,(n+1) \leq 2^n$$

The value of n is the least positive integer satisfying the inequality.

Putting $$n = 10$$

$$1100 \leq 2^{10} \quad \text{or} \quad 1100 \leq 1024$$

which is not true.

Putting $$n = 11$$

$$1200 \leq 2^{11} \quad \text{or} \quad 1200 \leq 2048$$

which is true.

$$\therefore \quad n = 11.$$

Ex. 10-12. *Show that if two symmetrical binomial distributions* $\left(p = q = \frac{1}{2}\right)$ *of degree n (and of the same number of observations) are so superposed that rth term of the one coincides with the $(r+1)$th term of the other, the distribution formed by adding superposed terms is a symmetrical binomial distribution of degree $(n+1)$.*

Sol. Let N be the number of observations. Then the successive terms of the binomial distribution are

$$N.\left(\frac{1}{2}\right)^n, N\,{}^nc_1\left(\frac{1}{2}\right)^n, \ldots\ldots\ldots N\,{}^nc_r\left(\frac{1}{2}\right)^n, \ldots\ldots\ldots N\,{}^nc_n\left(\frac{1}{2}\right)^n$$

$$\therefore \; r\text{th term of the first distribution} = N.{}^nc_{r-1}\frac{1}{2^n}$$

$$\text{and } (r+1)\text{th term of the second distribution} = N.{}^nc_r\frac{1}{2^n}$$

$$\therefore \text{Sum} = \frac{N}{2^n}\{{}^nc_{r-1} + {}^nc_r\}$$

$$= N.{}^{n+1}c_r \frac{1}{2^n}$$

$$= 2N.{}^{n+1}c_r \frac{1}{2^{n+1}}$$

which is the $(r+1)$ th term of the binomial distribution

$$2N.\left(\frac{1}{2}+\frac{1}{2}\right)^{n+1}$$

which is symmetrical binomial distribution of degree $(n+1)$ and total frequency $2N$.

Ex. 10-13. *Eight mice are selected at random and they are divided into two groups of 4 each. Each mouse in group A is given a dose of certain poison 'a' which is expected to kill one in four; each mouse in group B is given a dose of certain poison 'b' which is expected to kill one in two. Find the probability that the deaths in groups B are lesser than in group A.*

Sol. For group A,

$$n = 4,\ p = \frac{1}{4},\ q = \frac{3}{4}$$

$\therefore$ Prob. of x successes is given by

$$P(x) = {}^4c_x\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{4-x}$$

$$= \frac{{}^4c_x\,(3)^{4-x}}{256}$$

For group B,

$$n = 4,\ p = \frac{1}{2},\ q = \frac{1}{2}$$

$\therefore$ Prob. of x successes is given by

$$Q(x) = {}^4c_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x}$$

$$= \frac{{}^4c_x}{16}$$

$$\text{Reqd. prob.} = Q(0)\,\{P(1)+P(2)+P(3)+P(4)\}+Q(1)\{P(2) +P(3)+P(4)\}+Q(2)\,\{P(3)+P(4)\}+Q(3)P(4)$$

$$= \frac{1}{4096}\{{}^4c_0({}^4c_1.3^3 + {}^4c_2.3^2 + {}^4c_3.3 + {}^4c_4) +{}^4c_1.({}^4c_2.3^2 + {}^4c_3.3 + {}^4c_4) + {}^4c_2.({}^4c_3.3 + {}^4c_4) + {}^4c_3.({}^4c_4)\}$$

$$= \frac{1}{4096}\{(108+54+12+1)+4\,(54+12+1)+6(12+1)+4\}$$

Ex. 10.14. *Each of two perso*
obtain the same number of heads
Sol. For each person,

$\therefore$ Prob. of x heads is given b

P(

Since two persons toss indep
Reqd. prob. $= P(3)P(3) + P($

$= \left\{\left(\frac{1}{2}\right)\right.$

$= \frac{1}{64}\{1$

Ex. 10-15. *Find the probab*
failures to the probability of exac
Sol. Let p be the prob. of su
Then prob. of r failures $= {}^nc$
and prob. of $(n-r)$ failures ${}^nc_n$

$\therefore$ Ratio

This ratio can be independe

or

$\therefore$

Ex. 10-16. *Bring out the f*
binomial distribution is 5 and it
Sol. Here $np = 5$
and $\sqrt{}$
$\therefore$

$+1)$ and total frequency $2N$.

ey are divided into two groups of 4
oison 'a' which is expected to kill
ain poison 'b' which is expected to
oups B are lesser than in group A.

r

x

$1)\{P(2)$

$(3)P(4)$

$^3 + {}^4c_2 . 3^2 + {}^4c_3 . 3 + {}^4c_4)$

$+ {}^4c_4) + {}^4c_3 .({}^4c_4)\}$

$+1)+6(12+1)+4\}$

$$= \frac{525}{4096}.$$

Ex. 10.14. *Each of two persons tosses three fair coins. What is the probability that they obtain the same number of heads ?*

Sol. For each person,

$$n = 3$$

$$p = \text{prob. of head in a toss of a coin} = \frac{1}{2}$$

$$q = 1 - p = \frac{1}{2}$$

$\therefore$ Prob. of x heads is given by

$$P(x) = {}^n c_x p^x q^{n-x} = {}^3 c_x \left(\frac{1}{2}\right)^3$$

Since two persons toss independently,

Reqd. prob. $= P(3)P(3) + P(2)P(2) + P(1)P(1) + P(0)P(0)$

$$= \left\{\left(\frac{1}{2}\right)^3\right\}^2 + \left\{{}^3c_2\left(\frac{1}{2}\right)^3\right\}^2 + \left\{{}^3c_1\left(\frac{1}{2}\right)^3\right\}^2 + \left\{\left(\frac{1}{2}\right)^3\right\}^2$$

$$= \frac{1}{64}\{1 + 9 + 9 + 1\} = \frac{5}{16}.$$

Ex. 10-15. *Find the probability of success, if the ratio of the probability of exactly r failures to the probability of exactly (n–r) failures in n trials is independent of n.*

Sol. Let p be the prob. of success and $q = 1 - p$

Then prob. of r failures $= {}^n c_r q^r p^{n-r}$

and prob. of $(n - r)$ failures ${}^n c_{n-r}\, q^{n-r} p^r$

$\therefore$ Ratio $$= \left(\frac{p}{q}\right)^{n-2r}$$

This ratio can be independent of n only when

$$\frac{p}{q} = 1$$

or $$p = q = 1 - p$$

$\therefore$ $$p = \frac{1}{2}.$$

Ex. 10-16. *Bring out the fallacy, if any, in the following statement. The mean of a binomial distribution is 5 and its s.d. is 3.*

Sol. Here $np = 5$

and $$\sqrt{npq} = 3$$

$\therefore$ $$npq = 9$$

or $$5q = 9$$
$$q = 1.8$$

which is not true as probability is to be less than unity.

Ex. 10-17. *If on an average 1 vessel in every 10 is wrecked, find the probability that out of 5 vessels expected to arrive 4 at least will arrive safely.*

Sol. Let the arrival of a vessel safely be called success,

Then p = prob. of a vessel to arrive safely

= (1–prob. of a vessel to be wrecked)

$$= 1 - \frac{1}{10} = \frac{9}{10}$$

$$\therefore \quad q = \frac{1}{10}$$

Here $n = 5$

$$\therefore \quad P(x) = \frac{{}^5c_x . 9^x}{10^5}$$

Required prob. $= P(4) + P(5)$

$$= \frac{1}{10^5}\{{}^5c_4 . 9^4 + 9^5\}$$

$$= \frac{9^4.(5+9)}{10^5} = \frac{(14)9^4}{10^5} = 0{\cdot}91854.$$

Ex. 10-18. *'m' things are distributed among 'a' men and 'b' women, show that the chance that the number of things received by men is odd, is*

$$\frac{1}{2} \cdot \frac{(b+a)^m - (b-a)^m}{(b+a)^m}.$$

Sol. If one thing is distributed, prob. of a man to get

$$= \frac{a}{a+b}$$

and prob. for a woman to get it

$$= \frac{b}{a+b}$$

$\therefore$ Out of m things distributed, prob. for men to receive 'r' things

$$= {}^mc_r \left(\frac{a}{a+b}\right)^r \left(\frac{b}{a+b}\right)^{m-r}$$

$\therefore$ Prob. for men to receive odd number of things

$$= \sum {}^mc_r \left(\frac{a}{a+b}\right)^r \left(\frac{b}{a+b}\right)^{m-r}$$

where summation extends over odd values of r from 0 to m

$$= \frac{1}{(a+b)^m}\{{}^mc_1\, ab^{m-1} + {}^mc_3\, a^3 b^{m-3} + {}^mc_5\, a^5 b^{m-5} + \ldots\}$$

$$= \frac{1}{(a+b)^m}\left\{\frac{(b+a)^m - (b-}{2}\right.$$

Ex. 10-19. *Mean of a bino*
1·92. Find other constants of th

Sol. Here

$\therefore$ $q(q$

or $q(2q$

or $2q^2 - q - 0$

or $(2q + 0{\cdot}6)(q - ($

$\therefore$

$\therefore$

$\therefore$ From (i)

$\therefore$

Mode = greatest integer les
= 4

Variance = $\mu_2 = npq = 20($

$\therefore$ s.d. = $\sqrt{}$

Ex. 10-20. *The following da filter paper for 80 sets of seeds.*

x: 0 1 2 3
y: 6 20 28 12

Sol. Here $n = 10$, $N = 80$ an

$\therefore$ A

$$= \frac{1}{(a+b)^m}\left\{\frac{(b+a)^m-(b-a)^m}{2}\right\}$$

Ex. 10-19. *Mean of a binomial distribution is 4 and its third moment about mean is 1·92. Find other constants of the distribution.*

Sol. Here

$$np = 4 \qquad \ldots(i)$$

$$\mu_3 = npq\,(q-p) = 1\cdot 92 \qquad \ldots(ii)$$

$$\therefore \quad q(q-p) = 0\cdot 48$$

or $$q(2q-1) = 0\cdot 48$$

or $$2q^2 - q - 0\cdot 48 = 0$$

or $$(2q+0\cdot 6)(q-0\cdot 8) = 0$$

$$\therefore \quad q = 0\cdot 8 \text{ as } q \neq -0\cdot 3$$

$$\therefore \quad p = 0\cdot 2$$

$\therefore$ From (*i*)

$$n = \frac{4}{0\cdot 2} = 20$$

$$\therefore \quad n = 20,\ p = 0\cdot 2,\ q = 0\cdot 8$$

Mode = greatest integer less than $(n+1)\,p = 4\cdot 2$
$= 4$

Variance = $\mu_2 = npq = 20\,(0\cdot 2)\,(0\cdot 8) = 3\cdot 2$

$$\therefore \quad \text{s.d.} = \sqrt{\mu_2} = \sqrt{3\cdot 2}$$

$$\mu_4 = npq\,\{1+3pq(n-2)\}$$
$$= 3\cdot 2\,\{1+3(0\cdot 2)(0\cdot 8)\,18\}$$
$$= (3\cdot 2)(9\cdot 64) = 30\cdot 848$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(1\cdot 92)^2}{(3\cdot 2)^3} = 0\cdot 1125$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{30\cdot 848}{(3\cdot 2)^2} = 3\cdot 0125$$

$$\gamma_1 = \sqrt{\beta_1} = 0.3354$$

$$\gamma_2 = \beta_2 - 3 = 0\cdot 0125.$$

Ex. 10-20. *The following data are the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a binomial distribution to these data :*

x:	0	1	2	3	4	5	6	7	8	9	10	*Total*
y:	6	20	28	12	8	6	0	0	0	0	0	80

Sol. Here $n = 10$, $N = 80$ and $\Sigma f = 80$

$$\therefore \quad A.M. = \frac{\sum fx}{\sum f} = \frac{(20)1+(28)2+12(3)+(8)4+(6)5}{80}$$

$$= \frac{20+56+36+32+30}{80}$$

$$= \frac{174}{80}$$

$$\therefore \quad np = \frac{174}{80}$$

$$\therefore \quad p = \frac{1\cdot 74}{8} = 0\cdot 2175$$

$$q = 1 - p = 0\cdot 7825$$

Hence the binomial distribution to be fitted to the data is

$$80\,(0\cdot 7825 + 0\cdot 2175)^{10}$$

$\therefore$ Required binomial frequency distribution is

x:	0	1	2	3	4	5	6	7	8	9	10
y:	7	19	24	18	8	3	1	0	0	0	0

Ex. 10-21. *For a binomial variate x, find p if n = 4*

and $\quad P(x=4) = 6P(x=2).$

Sol. The distribution of x is

$$P(x) = {}^4c_x\, p^x q^{n-x}$$

Now $\quad P(4) = 6P(2)$

$$p^4 = 6\ {}^4c_2\, p^2 q^2$$

i.e., $\quad p^2 = 36q^2 \qquad$ (assuming $p \neq 0$)

$\Rightarrow \quad p = 6q \qquad (\because p, q \nless 0)$

$\Rightarrow \quad = 6 - 6p$

$$p = \frac{6}{7}.$$

Ex. 10-22. *For a binomial distribution the mean is 4 and variance is 2. Find the distribution.*

Sol. $np = 4, \qquad npq = 2$

$$\therefore \quad q = \frac{1}{2}$$

$$\therefore \quad p = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\therefore \quad n = 8$$

$\therefore$ Binomial distribution is

$$P(x) = 8c_x \left(\frac{1}{2}\right)^8, \; x = 0, 1, \ldots\ldots 8.$$

Ex. 10-23. *If x and y are binom*

Find P $(x+y \geq 1)$

Sol. Since probabilities of succ

$n = 10 + 5 = 15, \; p = \frac{1}{2}.$

Let

Then distribution of z is

$P(z$

$\therefore \quad P(z \geq 1$

Ex. 10-24. *Starting with the id*

$\sum_{x=0}^{n}$

find mean and variance of B.D

Sol. By given identity

$\sum_{x=0}^{n}$

Differentiating w.r.t. p

$\sum_{x=0}^{n} {}^nc_x \{xp^{x-1}q^{n-x} - (n-x)q^n$

i.e., $\quad \sum_{x=0}^{n} {}^nc_x p^{x-1} q^{n-x-}$

$\Rightarrow \quad \sum_{x=0}^{n} {}^nc_x p^x q^{n-x}(x$

i.e., $\quad \sum_{x=0}^{n} x\,{}^nc_x p^x q^{n-x}$ -

i.e., $\quad \bar{x} - np(q+p)^n = ($

$\Rightarrow \quad \bar{x} = np$

s

7	8	9	10
0	0	0	0

(assuming $p \neq 0$)

($\because p, q \nless 0$)

4 and variance is 2. Find the

.8.

Ex. 10-23. *If x and y are binomial variates with $n=10, p=\frac{1}{2}$ and $n=5, p=\frac{1}{2}$ respectively. Find $P(x+y \geq 1)$*

Sol. Since probabilities of success for x and y are same, $x+y$ is a binomial variate with $n = 10+5 = 15,\ p = \frac{1}{2}$.

Let $\qquad z = x + y$

Then distribution of z is

$$P(z) = {}^{15}c_z\left(\frac{1}{2}\right)^{15},\ z = 0, 1, \ldots\ldots 15$$

$$\therefore \qquad P(z \geq 1) = 1 - P(z < 1)$$
$$= 1 - P(z = 0)$$
$$= 1 - \left(\frac{1}{2}\right)^{15}$$

Ex. 10-24. *Starting with the identity*

$$\sum_{x=0}^{n} {}^{n}c_x\, p^x q^{n-x} = (q+p)^n$$

find mean and variance of B.D.

Sol. By given identity

$$\sum_{x=0}^{n} {}^{n}c_x\, p^x q^{n-x} = (q+p)^n$$

Differentiating w.r.t. p

$$\sum_{x=0}^{n} {}^{n}c_x \{xp^{x-1}q^{n-x} - (n-x)q^{n-x-1}p^x\} = n(q+p)^{n-1}(-1+1) \qquad \left(\because \frac{dq}{dp} = -1\right)$$

i.e., $$\sum_{x=0}^{n} {}^{n}c_x p^{x-1} q^{n-x-1}\{xq - (n-x)p\} = 0$$

$\Rightarrow$ $$\sum_{x=0}^{n} {}^{n}c_x p^x q^{n-x}(x - np) = 0 \qquad \ldots(1)$$

i.e., $$\sum_{x=0}^{n} x\,{}^{n}c_x p^x q^{n-x} - np\sum_{x=0}^{n} {}^{n}c_x p^x q^{n-x} = 0$$

i.e., $$\bar{x} - np(q+p)^n = 0$$

$\Rightarrow$ $$\bar{x} = np$$

Differentiating (1) w.r.t. p

$$\sum_{x=0}^{n} {}^{n}c_x [\{xp^{x-1}q^{n-x} - (n-x)p^x q^{n-x-1}\}(x-np) + p^x q^{n-x}(-n)] = 0$$

i.e., $$\sum_{x=0}^{n} {}^{n}c_x p^{x-1}q^{n-x-1}\{xq-(n-x)p\}(x-np) - n\sum_{x=0}^{n} {}^{n}c_x p^x q^{n-x} = 0$$

$$\Rightarrow \quad \frac{1}{p.q}\sum_{x=0}^{n} {}^{n}c_x p^x q^{n-x}(x-np)^2 - n(q+p)^n = 0$$

$\Rightarrow$ $\mu_2 = npq.$

Ex. 10-25. *Two dice are thrown n times. Let x denotes the number of throws in which the number on first die exceeds the number on the second die. Find the distribution of x.*

Sol. Let in a throw success means :

"no. on first die exceeds the no. on second die".

Different possibilities of success are :

(2, 1), (3, 1), (3, 2), (4, 1), (4, 2), (4, 3)

(5, 1), (5, 2), (5, 3), (5, 4), (6, 1), (6, 2), (6, 3) (6, 4), (6, 5)

$\therefore$ p = prob. of success in one throw $= \dfrac{15}{36} = \dfrac{5}{12}$

$\therefore$ x is a B,V with no. of trials $= n$ and prob. of success $= \dfrac{5}{12}$.

$$\therefore \qquad x \sim b\left(n, \frac{5}{12}\right)$$

Ex. 10-26. *If x is binomially distributed with parameters n and p and y is beta distributed with parameters k and $n-k+1$, then*

$$P(x \ge k) = \frac{1}{\beta(k, n-k+1)}\int_0^p u^{k-1}(1-u)^{n-k}\,du$$

and hence $F_y(p) = 1 - F_x(k-1)$. {where $F_x(.)$ denotes the $c.d.f$ of x etc.}

Sol. Let
$$P = P(x \ge k)$$
$$= 1 - P(x<k)$$
$$= 1 - \sum_{j=0}^{k-1} {}^{n}c_j p^j q^{n-j}$$
$$= \cancel{1} q^n - \sum_{j=1}^{k-1} {}^{n}c_j p^j q^{n-j}$$
$$\frac{dP}{dp} = nq^{n-1} - \sum_{j=1}^{k-1} {}^{n}c_j\left[j p^{j-1}q^{n-j} - (n-j)p^j q^{n-j-1}\right]$$
$$= nq^{n-1} - \sum_{j=1}^{k-1}\left\{\left(j.{}^{n}c_j\right)p^{j-1}q^{n-j} - (n-j)\,{}^{n}c_j\,p^j q^{n-j-1}\right\}$$

$$= nq^{n-1} - \sum_{j=1}^{k-1}\left\{\frac{j\,n!}{(j)!(n-j)!}\right.$$

$$= nq^{n-1} - \sum_{j=1}^{k-1}\left\{\frac{n!}{(j-1)!(n-j}\right.$$

$$= nq^{n-1} - \sum_{j=1}^{k-1}\{n.{}^{n-1}c_{j-1}\,p^{j-}$$

$$= nq^{n-1} - n\sum_{j=1}^{k-1}\{{}^{n-1}c_{j-1}\,p^{j-1}$$

$$= nq^{n-1} - n\left[{}^{n-1}c_0\,q^{n-1} - {}^{n-1}\right.$$

$$= n.{}^{n-1}c_{k-1}\,p^{k-1}q^{n-k}$$

Integrate w.r.t. p

$$[P]^p_{p=0} = n.{}^{n-1}c_{k-1}\int_0^p p^{k-}$$

$$= n.{}^{n-1}c_{k-1}\int_0^p u^{k-1}$$

$$P(x \ge k) = n.\frac{(n-1)!}{(k-1)!(n-k)!}$$

$$= \frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)}$$

$$= \frac{1}{\beta(k, n-k+1)}\int_0^p$$

$\therefore$ $1 - P(x<k) = F_y(p)$

i.e., $1 - F_x(k-1) = F_y(p)$

Ex. 10-27. *A drunk performs a He starts at zero. He takes succ p and to the left with probability 1*

after n steps. Find the distribution

Sol. Let x_i denote the ith step

Then x

Define a variate $y_i s.t.\ y_i = \dfrac{x_i}{2}$

$$= nq^{n-1} - \sum_{j=1}^{k-1}\left\{\frac{j\,n!}{(j)!\,(n-j)!}\,p^{j-1}\,q^{n-j} - (n-j)\frac{n!}{j!(n-j)!}\,p^{j}\,q^{n-j-1}\right\}$$

$$= nq^{n-1} - \sum_{j=1}^{k-1}\left\{\frac{n!}{(j-1)!\,(n-j)!}\,p^{j-1}\,q^{n-j} - \frac{n!}{j!(n-j-1)!}\,p^{j}q^{n-j-1}\right\}$$

$$= nq^{n-1} - \sum_{j=1}^{k-1}\left\{n.\,{}^{n-1}c_{j-1}\,p^{j-1}\,q^{n-j} - n.\,{}^{n-1}c_j\,p^j q^{n-1-j}\right\}$$

$$= nq^{n-1} - n\sum_{j=1}^{k-1}\left\{{}^{n-1}c_{j-1}\,p^{j-1}q^{\overline{n-1}-\overline{j-1}} - {}^{n-1}c_j\,p^j\,q^{n-1-j}\right\}$$

$$= nq^{n-1} - n\left[{}^{n-1}c_0\,q^{n-1} - {}^{n-1}c_{k-1}\,p^{k-1}\,q^{n-k}\right]$$

$$= n.\,{}^{n-1}c_{k-1}\,p^{k-1}\,q^{n-k}$$

Integrate w.r.t. p

$$[P]^{p}{}_{p=0} = n.\,{}^{n-1}c_{k-1}\int_o^p p^{k-1}(1-p)^{n-k}\,dp.$$

$$= n.\,{}^{n-1}c_{k-1}\int_o^p u^{k-1}(1-u)^{n-k}\,du$$

$$P(x \geq k) = n.\frac{(n-1)!}{(k-1)!\,(n-k)!}\int_o^p u^{k-1}(1-u)^{n-k}\,du$$

$$= \frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)}\int_o^p u^{k-1}(1-u)^{n-k}\,du$$

$$= \frac{1}{\beta(k, n-k+1)}\int_o^p u^{k-1}\,(1-u)^{n-k}\,du.$$

$\therefore\; 1 - P(x < k) = F_y(p)$

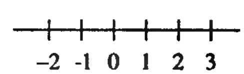
i.e., $1 - F_x(k-1) = F_y(p)$

Ex. 10-27. *A drunk performs a 'random walk' over positions* $0, \pm 1, \pm 2, \ldots\ldots$ *as follows : He starts at zero. He takes successive one-unit steps, going to the right with probability p and to the left with probability 1-p. His steps are independent. Let x denote his position after n steps. Find the distribution of* $\dfrac{x+n}{2}$ *and then find* $E(x)$.

Sol. Let x_i denote the ith step of the drunk.

Then $\quad x_i = 1$ if drunk goes to right

$\quad\quad = -1$ if drunk goes to left



Define a variate y_i s.t. $y_i = \dfrac{x_i + 1}{2}$

Then $y_i = 1$ if $x_i = 1$

$= 0$ if $x_i = -1$

Then $P(y_i = 1) = p$ and $P(y_i = 0) = 1 - p = q$.

Let
$$z = \sum_{i=1}^{n} y_i = \frac{\Sigma(x_i + 1)}{2} = \frac{\Sigma x_i + n}{2} = \frac{x + n}{2}$$

Then z is a B.V with parameters n, p.

Also
$$E(z) = \frac{E(x) + n}{2} = np$$

$\therefore$
$$E(x) = 2np - n$$
$$= n\{2p - 1\}.$$

Ex. 10-28. *If x has a $b(n, p)$ and y follows negative binomial distribution with parameters r and p.*
Show that
$$F_x(r-1) = 1 - F_y(n-r)$$

Sol.
$$\text{R.H.S.} = 1 - F_y(n-r)$$
$$= 1 - P(Y \le n-r)$$
$$= P(Y > n-r)$$
$$= \sum_{Y=n-r+1}^{\infty} {}^{y+r-1}c_{r-1}\, q^y p^r$$

Put
$$y = (n - r + 1) + t$$

$\therefore$
$$\text{R.H.S.} = \sum_{t=0}^{\infty} {}^{n+t}c_{r-1}\, q^{n-r+1+t}\, p^r$$
$$= p^r q^{n-r+1} \sum_{t=0}^{\infty} {}^{n+t}c_{r-1}\, q^t$$
$$= p^r q^{n-r+1} \sum_{t=0}^{\infty} \left\{ \sum_{k=0}^{r-1} {}^{n}c_k \cdot {}^{t}c_{r-1-k} \right\} q^t$$
$$\left\{ \because \sum_{k=0}^{r-1} {}^{n}c_k\, {}^{t}c_{r-1-k} = {}^{n+t}c_{r-1} \right\}$$
$$= p^r q^{n-r+1} \sum_{k=0}^{r-1} {}^{n}c_k \left\{ \sum_{t=0}^{\infty} {}^{t}c_{r-1-k}\, q^t \right\}$$

Take $t - (r - 1 - k) = j$

$= p^r q^n$

$= p^r q^n$

$= p^r q^n$

$= p^r q^n$

$= p^r q^n$

$= \sum_{k=0}^{r-1} {}^{n}$

Ex. 10-29. *Let $x_1; x_2$ be t*

$p_1 < p_2$, show that

$P(x_1 \le$

Sol. Let x be a B.V. with par

$P = p(x \le k) = \sum_{x=0}^{k} {}^{n}$

$= q^n + {}^{n}$

$+ {}^{n}c_k$

$\frac{dP}{dp} = -nq^{n-1} + {}^{n}c_1 \{q^{n-1} -$

$+ \ldots + {}^{n}c_{k-1}\{(k-1)p^{k-2} q^{n-k}$

$+ {}^{n}c_k\{kp^{k-1} q^{n-k} - (n-k)p^k$

$$= p^r q^{n-r+1} \sum_{k=0}^{r-1} {}^n c_k \left\{ \sum_{t=r-1-k}^{\infty} {}^t c_{r-1-k}\, q^t \right\}$$

$\left(\because n.\ {}^t c_j = \ldots \text{ if } j > \ldots\right.$

Take $t-(r-1-k)=j$

$$= p^r q^{n-r+1} \sum_{k=0}^{r-1} {}^n c_k \left\{ \sum_{j=0}^{\infty} {}^{r-1-k+j} c_{r-1-k}\, q^{j+r-1-k} \right\}$$

$$= p^r q^n \sum_{k=0}^{r-1} {}^n c_k\, q^{-k} \left(\sum_{j=0}^{\infty} {}^{r-1-k+j} c_j\, q^j \right)$$

$$= p^r q^n \sum_{k=0}^{r-1} {}^n c_k\, q^{-k} \left\{ 1 + {}^{r-k} c_1\, q + {}^{r-k+1} c_2\, q^2 \ldots \right\}$$

$$= p^r q^n \sum_{k=0}^{r-1} {}^n c_k\, q^{-k} \left\{ 1 + (r-k)q + \frac{(r-k+1)(r-k)}{2!} q^2 \ldots \right\}$$

$$= p^r q^n \sum_{k=0}^{r-1} {}^n c_k\, q^{-k} (1-q)^{-(r-k)}$$

$$= \sum_{k=0}^{r-1} {}^n c_k\, p^k q^{n-k} = F_x(r-1).$$

Ex. 10-29. *Let $x_1; x_2$ be two B.Vs with parameters $n_1, p_1; n_2, p_2$ respectively. If $p_1 < p_2$, show that*

$$P(x_1 \le k) \ge P(x_2 \le k), (k = 0, 1, \ldots\ldots n).$$

Sol. Let x be a B.V. with parameters n, p

$$P = p(x \le k) = \sum_{x=0}^{k} {}^n c_x\, p^x q^{n-x}$$

$$= q^n + {}^n c_1\, p q^{n-1} + {}^n c_2\, p^2 q^{n-2} + \ldots + {}^n c_{k-1}\, p^{k-1} q^{n-k+1} + {}^n c_k\, p^k q^{n-k}$$

$$\frac{dP}{dp} = -nq^{n-1} + {}^n c_1 \left\{ q^{n-1} - (n-1) p q^{n-2} \right\} + {}^n c_2 \left\{ 2pq^{n-2} - p^2 (n-2) q^{n-3} \right\}$$

$$+ \ldots\ldots + {}^n c_{k-1} \left\{ (k-1) p^{k-2} q^{n-k+1} - (n-k+1) p^{k-1} q^{n-k} \right\}$$

$$+ {}^n c_k \left\{ k p^{k-1} q^{n-k} - (n-k) p^k q^{n-k-1} \right\}$$

$$= -(n-k)\,{}^nc_k\,p^k\,q^{n-k-1}$$

$$< 0$$

$\therefore$ P decreases as p increases

$\Rightarrow \therefore \; P(x_1 \le k) \ge P(x_2 \ge k).$

EXERCISES

1. If x is a $b(n, p)$ {*i.e.*, binomial variate with parameters n, p}, then show that

$$E\left(\frac{x}{n}\right) = p$$

and
$$E\left\{\frac{x}{n} - p\right\}^2 = \frac{p(1-p)}{n}.$$

2. If x is a random variable distributed according to the binomial law $P(x = k) \equiv b(k) = {}^nc_k p^k q^{n-k}, k = 0, 1, \ldots\ldots, q = 1 - p,$ show that

$$\frac{b(k+1)}{b(k)} = \frac{n-k}{k+1} \cdot \frac{p}{k}.$$

3. If x is binomially distributed with parameters n and p, what is the distribution of $y = n - x$.

4. The *m.g.f.* of a random variable x is $\left(\frac{2}{3} + \frac{1}{3}e^t\right)^9$. Show that

$$P\{\mu - 2\sigma < x < \mu + 2\sigma\} = \sum_{x=1}^{5} {}^9c_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}$$

5. A's chance of winning a game against B is $\frac{2}{3}$. Find his chance of winning at least three games out of five.

$$\left[\textbf{Ans. } \frac{192}{243}\right]$$

6. The incidence of occupational disease in an industry is such that the workers have a 20% chance of suffering from it. What is the probability that out of six workers 4 or more will catch the disease.

$$\left[\textbf{Ans. } \frac{53}{3125}\right]$$

7. If 5 coins are tossed, what is the probability that there shall be at least 4 heads ?

$$\left[\textbf{Ans. } \frac{3}{16}\right]$$

8. Out of 4000 families with 4 children each, how many would you expect to have at least 1 boy ? Assume that the probability of a male birth is $\frac{1}{2}$. **[Ans.** 3750]

9. Find the probability of gue
examination.

10. If we take 100 sets of 10 to:
to get 7 heads and 3 tails ?

11. In the above example in ho

12. An ordinary six-sided die i
4, 3, 2, 0 aces ?

13. An experiment succeeds tw
trials there will be at least 4

14. A teacher claims that he co
whether they will obtain I
demonstrate his claim he fo
being correct in 4 cases.

15. In litters of 4 mice the nu
noted. The figures are give
No. of female mice 0
No. of litters 9
If the chance of obtaining a
constant of unknown proba
[Ans. 0·466, exp

16. Ten coins are tossed 1024
these frequencies with the
the data.
No. of heads (x) 0
Frequencies (f) 3
No. of heads (x) 7
Frequencies (f) 128

[**Ans.** x :
f :

17. Out of 800 families with 4 c
(*i*) 2 boys and 2 girls.
(*ii*) at least one boy.
(*iii*) no girl.
(*iv*) at most 2 girls ?

$n, p\}$, then show that

inomial law $P(x = k) \equiv b(k) =$

p, what is the distribution of

w that

his chance of winning at least

$\left[\textbf{Ans. } \frac{192}{243}\right]$

is such that the workers have a
lity that out of six workers 4 or

$\left[\textbf{Ans. } \frac{53}{3125}\right]$

e shall be at least 4 heads ?

$\left[\textbf{Ans. } \frac{3}{16}\right]$

ny would you expect to have at

irth is $\frac{1}{2}$. **[Ans.** 3750]

9. Find the probability of guessing correctly at least 6 of the 10 answers on a true-false examination.

$$\left[\textbf{Ans. } \frac{193}{512}\right]$$

10. If we take 100 sets of 10 tosses of a perfect coin, in how many cases should we expect to get 7 heads and 3 tails ? **[Ans.** 12]

11. In the above example in how many cases should we expect to get 7 heads at least? **[Ans.** 17]

12. An ordinary six-sided die is thrown 4 times. What are the probabilities of obtaining 4, 3, 2, 0 aces ?

$$\left[\textbf{Ans. } \frac{625}{1296}, \frac{25}{216}, \frac{5}{324}, \frac{1}{1296}\right]$$

13. An experiment succeeds twice as often as it fails. Find the chance that in the next six trials there will be at least 4 successes.

$$\left[\textbf{Ans. } \frac{496}{729}\right]$$

14. A teacher claims that he could often tell while his students were still in their first year whether they will obtain I, II, III divisions or fail in their final examinations. To demonstrate his claim he forecasts the fates of 8 students. Find the probability of his being correct in 4 cases.

$$\left[\textbf{Ans. } \frac{2835}{32768}\right]$$

15. In litters of 4 mice the number of litters which contained 0, 1, 2, 3, 4 females were noted. The figures are given in the table below :

No. of female mice	0	1	2	3	4	Total
No. of litters	9	30	35	24	5	103

If the chance of obtaining a female in a single trial is assumed constant, estimate this constant of unknown probability. Find also expected frequencies.

[Ans. 0·466, expected frequencies are the respective terms in the binomial expansion of $103\,(0{\cdot}534 + 0{\cdot}466)^4$]

16. Ten coins are tossed 1024 times and the following frequencies observed. Compare these frequencies with the expected frequencies obtained by fitting binomial dist. to the data.

No. of heads (x)	0	1	2	3	4	5	6
Frequencies (f)	3	8	39	106	188	257	226
No. of heads (x)	7	8	9	10			
Frequencies (f)	128	59	7	3			

Ans.

x:	0	1	2	3	4	5	6	7	8	9	10
f:	1	10	45	120	210	252	210	120	45	10	1

17. Out of 800 families with 4 children each, how many families would you expect to have

(*i*) 2 boys and 2 girls.

(*ii*) at least one boy.

(*iii*) no girl.

(*iv*) at most 2 girls ?

Assume equal probabilities for boys and girls.

[Ans. (*i*) 300, (*ii*) 750 (*iii*) 50, (*iv*) 550]

18. A room has three lamp sockets. From a collection of 10 light bulbs, of which only 6 are good, three are selected at random and are put in the sockets. What is the probability that there shall be light ?

10-2. Poisson Distribution (P.D.)

Poisson Probability Distribution. *The poisson probability dist. of the variate x is*

$$P(x) = e^{-m}\frac{m^x}{x!}, x = 0, 1, \ldots \infty$$

The variate x is called **Poisson Variate** *and m is called the parameter of the distribution.*

Poisson Frequency Distribution. *The poisson frequency dist. of the variate x is*

$$F(x) = Ne^{-m}\frac{m^x}{x!}, x = 0, 1, \ldots \infty$$

where N is the total frequency.

Poisson distribution is regarded as the limiting form of binomial distribution when n (no. of trials) approaches '∞' and p (prob. of success) approaches zero such that np remains a finite constant m.

For B.D. $$P(x) = {}^nc_x\, p^x q^{n-x}$$

$$\underset{\substack{n\to\infty \\ p\to 0 \\ np=m}}{\text{Lt}} P(x) = \underset{\substack{n\to\infty \\ p\to 0 \\ np=m}}{\text{Lt}} {}^nc_x p^x q^{n-x}$$

$$= \underset{n\to\infty}{\text{Lt}} \frac{n!}{x!(n-x)!}\left(\frac{m}{n}\right)^x \left(1-\frac{m}{n}\right)^n$$

$$= \underset{n\to\infty}{\text{Lt}} \frac{n!}{x!(n-x)!}\left(\frac{m}{n}\right)^x \left(1-\frac{m}{n}\right)^n \left(1-\frac{m}{n}\right)^{-x}$$

$$= \frac{m^x}{x!}\underset{n\to\infty}{\text{Lt}} \frac{n(n-1)\ldots(n-x+1)}{n^x}\left(1-\frac{m}{n}\right)^n \left(1-\frac{m}{n}\right)^{-x}$$

$$= \frac{m^x}{x!}\underset{n\to\infty}{\text{Lt}} \left(\frac{n}{n}\right).\left(1-\frac{1}{n}\right)\ldots\ldots$$

$$\ldots\ldots\left(1-\frac{x-1}{n}\right)\left[\left(1-\frac{m}{n}\right)^{-n/m}\right]^{-m}\left(1-\frac{m}{n}\right)^{-x}$$

$$= \frac{m^x}{x!}e^{-m} \qquad \left\{\because \underset{n\to\infty}{\text{Lt}}\left(1-\frac{m}{n}\right)^{-n/m} = e\right\}$$

$\therefore$ Probability of x successes for Poisson distribution

$$= \frac{m^x}{x!}e^{-m}$$

i) 300, (ii) 750 (iii) 50, (iv) 550]
f 10 light bulbs, of which only 6
e sockets. What is the probability

ability dist. of the variate x is

.∞

the parameter of the distribution.
ency dist. of the variate x is

,...∞

of binomial distribution when n
oaches zero such that np remains

$$\Big)^x \left(1-\frac{m}{n}\right)^n$$

$$\Big)^x \left(1-\frac{m}{n}\right)^n \left(1-\frac{m}{n}\right)^{-x}$$

$$\frac{(n-x+1)}{^x}\left(1-\frac{m}{n}\right)^n \left(1-\frac{m}{n}\right)^{-x}$$

$$\frac{1}{n}\Big)\ldots\ldots$$

$$\frac{-1}{n}\Big)\left[\left(1-\frac{m}{n}\right)^{-n/m}\right]^{-m} \left(1-\frac{m}{n}\right)^{-x}$$

$$\left.\underset{n\to\infty}{\text{Lt}}\left(1-\frac{m}{n}\right)^{-n/m} = e\right\}$$

10.2.1. First four moments about mean

For Poisson distribution the probability of x successes is given by

$$P(x) = e^{-m}\frac{m^x}{x!},\ x = 0, 1, 2, \ldots\ldots.\infty$$

$$\mu_1'(0) = \sum_{x=0}^{n} xP(x) = \sum_{x=0}^{\infty} xe^{-m}\frac{m^x}{x!}$$

$$= e^{-m}\left\{1.m + 2\frac{m^2}{2!} + 3.\frac{m^3}{3!} + \ldots\ldots.\right\}$$

$$= me^{-m}\left\{1 + m + \frac{m^2}{2!} + \ldots\ldots..\right\}$$

$$= me^{-m}.e^{m} = m$$

$$\mu_2'(0) = \sum_{x=0}^{\infty} x^2 P(x) = \sum_{x=0}^{\infty}\{x(x-1)+x\}P(x)$$

$$= e^{-m}\sum_{x=0}^{\infty} x(x-1)\frac{m^x}{x!} + \sum_{x=0}^{\infty} xP(x)$$

$$= e^{-m}\left\{2\cdot 1.\frac{m^2}{2!} + 3\cdot 2.\frac{m^3}{3!} + \ldots\ldots.\right\} + m$$

$$= e^{-m}m^2(1+m+\ldots) + m$$

$$= e^{-m}m^2e^{m} + m = m^2 + m$$

$$\therefore \quad \mu_2 = \mu_2'(0) - \{\mu_1'(0)\}^2 = m^2 + m - m^2 = m$$

$$\mu_3'(0) = \sum_{x=0}^{\infty} x^3 P(x)$$

Writing x^3 as $x(x-1)(x-2) + 3x(x-1) + x$

$$\mu_3'(0) = \sum_{x=0}^{\infty}\{x(x-1)(x-2) + 3x(x-1) + x\}P(x)$$

$$= m^3 + 3m^2 + m$$

$$\therefore \quad \mu_3 = \mu_3'(0) - 3\mu_2'(0)\,\mu_1'(0) + 2\{\mu_1'(0)\}^3$$

$$= m^3 + 3m^2 + m - 3(m^2+m)m + 2m^3 = m$$

$$\mu_4'(0) = \sum_{x=0}^{\infty} x^4\,P(x)$$

But $x^4 \equiv x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x$

$$\mu_4'(0) = \sum_{x=0}^{\infty}\{x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x\}P(x)$$

$$= m^4+6m^3+7m^2+m.$$

$\therefore\ \mu_4' = \mu_4'(0)-4\mu_3'(0)\mu_1'(0)+6\mu_2'(0)\{\mu_1'(0)\}^2-3\{\mu_1'(0)\}^4$

$= m^4+6m^3+7m^2+m-4(m^3+3m^2+m)m+6(m^2+m)m^2-3m^4$

$= 3m^2+m.$

Ex. 10-30. *In a Poisson distribution, P(x) fox x = 0 is 10%. Find the mean.*

Sol. Let m be the mean

Then $$P(x) = e^{-m}\frac{m^x}{x!}$$

$\therefore$ $$P(0) = e^{-m}$$

$\therefore$ $$e^{-m} = 0\cdot1$$

$\therefore$ $$e^m = \frac{1}{0\cdot1} = 10$$

$\therefore$ $$m = \log_e 10 = 2\cdot3026.$$

Ex. 10-31. *Deduce first four moments about mean for Poisson distribution from those of binomial distribution.*

Sol. Poisson distribution is the limiting form of binomial distribution when n (no. of trials) tends to infinity and p (prob. of success) tends to zero such that np remains a finite constant m.

For binomial distribution,

$$\mu_1'(0) = np,\ \mu_2 = npq,\ \mu_3 = npq(q-p)$$

and $$\mu_4 = npq\{1+3pq(n-2)\}$$

$\therefore$ $$\mu_1'(0) \text{ for } P.D. = \mathop{\text{Lt}}_{\substack{n\to\infty\\ p\to0\\ np=m}} \mu_1'(0) \text{ for } B.D.$$

$$= \mathop{\text{Lt}}_{\substack{n\to\infty\\ p\to0\\ np=m}} (np) = m$$

$$\mu_2 \text{ for } P.D. = \mathop{\text{Lt}}_{\substack{n\to\infty\\ p\to0\\ np=m}} \mu_2 \text{ for } B.D.$$

$$= \mathop{\text{Lt}}_{\substack{n\to\infty\\ p\to0\\ np=m}} (npq)$$

$$= \mathop{\text{Lt}}_{n\to\infty} m\left(1-\frac{m}{n}\right) = m$$

μ_3 for

and μ_4 for

Ex. 10-32. *Let λ and μ_r*
distribution respectively. Obtai

Hence deduce the values o

Sol. By def.,

$\therefore \dfrac{d\mu_r}{d\lambda} = \sum_{x=0}^{n}\dfrac{1}{x!}\{-e^{-\lambda}(x-$

$= \sum_{x=0}^{n}\dfrac{1}{x!}\{e^{-\lambda}\lambda^{x-1}(x-\lambda)^r$

$= \dfrac{1}{\lambda}\sum_{x=0}^{\infty}(x-\lambda)^{r+1}e^{-\lambda}\dfrac{\lambda^x}{x!}-$

$= \dfrac{1}{\lambda}\mu_{r+1}-r\mu_{r-1}$

$$\mu_3 \text{ for } P.D. = \underset{\substack{n\to\infty \\ p\to 0 \\ np=m}}{\text{Lt}} \mu_3 \text{ for } B.D.$$

$$= \underset{\substack{n\to\infty \\ p\to 0 \\ np=m}}{\text{Lt}} npq(q-p)$$

$$= \underset{n\to\infty}{\text{Lt}}\ m\left(1-\frac{m}{n}\right)\left(1-\frac{2m}{n}\right) = m$$

and

$$\mu_4 \text{ for } P.D. = \underset{\substack{n\to\infty \\ p\to 0 \\ np=m}}{\text{Lt}} \mu_4 \text{ for } B.D.$$

$$= \underset{\substack{n\to\infty \\ p\to 0 \\ np=m}}{\text{Lt}} npq\{1+3pq(n-2)\}$$

$$= \underset{n\to\infty}{\text{Lt}}\ m\left(1-\frac{m}{n}\right)\left\{1+\frac{3m}{n}\left(1-\frac{m}{n}\right)(n-2)\right\}$$

$$= \underset{n\to\infty}{\text{Lt}}\ m\left(1-\frac{m}{n}\right)\left\{1+3m\left(1-\frac{m}{n}\right)\left(1-\frac{2}{n}\right)\right\}$$

$$= 3m^2 + m.$$

Ex. 10-32. *Let λ and μ_r denote the mean and central rth moment of a Poisson distribution respectively. Obtain the recurrence formula*

$$\mu_{r+1} = r\lambda\mu_{r-1} + \frac{d\mu_r}{d\lambda}$$

Hence deduce the values of β_1 and β_2.

Sol. By def.,

$$\mu_r = \sum_{x=0}^{\infty}(x-\lambda)^r e^{-\lambda}\frac{\lambda^x}{x!}$$

$$\therefore \frac{d\mu_r}{d\lambda} = \sum_{x=0}^{n}\frac{1}{x!}\{-e^{-\lambda}(x-\lambda)^r\lambda^x + x\lambda^{x-1}e^{-\lambda}(x-\lambda)^r - r(x-\lambda)^{r-1}e^{-\lambda}.\lambda^x\}$$

$$= \sum_{x=0}^{n}\frac{1}{x!}\{e^{-\lambda}\lambda^{x-1}(x-\lambda)^r(x-\lambda) - r(x-\lambda)^{r-1}e^{-\lambda}.\lambda^x\}$$

$$= \frac{1}{\lambda}\sum_{x=0}^{\infty}(x-\lambda)^{r+1}e^{-\lambda}\frac{\lambda^x}{x!} - r\sum_{x=0}^{\infty}(x-\lambda)^{r-1}e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= \frac{1}{\lambda}\mu_{r+1} - r\mu_{r-1}$$

$$\therefore \quad \mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}$$

Put $\quad r = 1, 2$ and 3

$$\mu_2 = \lambda\mu_0 + \lambda\frac{d\mu_1}{d\lambda} = \lambda \text{ as } \mu_0 = 1 \text{ and } \mu_1 = 0$$

$$\therefore \quad \mu_3 = 2\lambda\mu_1 + \lambda\frac{d\mu_2}{d\lambda} = \lambda$$

and

$$\mu_4 = 3\lambda\mu_2 + \lambda\frac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda$$

$$\therefore \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{\lambda} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda}.$$

Ex. 10-33. *For a Poisson variate x with parameter λ, show that*

$$\mu'_{r+1} = \lambda\mu'_r + \lambda\frac{d\mu'_r}{d\lambda}$$

where $\mu'_r = E(x^r)$ *and r is a non-negative integer.*

Sol. By def.,

$$\mu'_r = E(x^r)$$

$$= \sum_{x=0}^{\infty} x^r e^{-\lambda}\frac{\lambda^x}{x!}$$

$$\frac{d\mu'_r}{d\lambda} = \sum_{x=0}^{\infty}\frac{x^r}{x!}\{-e^{-\lambda}\lambda^x + e^{-\lambda}.x\lambda^{x-1}\}$$

$$= -\sum_{x=0}^{\infty} x^r e^{-\lambda}\frac{\lambda^x}{x!} + \frac{1}{\lambda}.\sum_{x=0}^{\infty} x^{r+1}e^{-\lambda}.\frac{\lambda^x}{x!}$$

$$= -\mu'_r + \frac{1}{\lambda}\mu'_{r+1}$$

$$\Rightarrow \quad \mu'_{r+1} = \lambda\mu'_r + \lambda\frac{d\mu'_r}{d\lambda}$$

10.2.2. Measure of skewness and kurtesis

By def.

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \sqrt{\frac{m^2}{m^3}} = \frac{1}{\sqrt{m}}$$

$$\therefore \quad \gamma_1 \to 0 \text{ as } m \to \infty$$

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{3m^2 + m}{m^2} - 3 = \frac{1}{m}$$

$$\gamma_2 \to 0 \text{ as } m \to \infty.$$

Ex. 10-34. *For Poisson dist*

Sol. L.I

10.2.3. Mode of the Poisson Dis

In Poisson distribution the p

P

The mode is that value of x

$P(x+1)$ *i.e.*,

Consider

or e^{-m}

or

Similarly other inequality gi

From (*i*) and (*ii*) modal valu

Case I. If m is an integer, the

Now $\frac{P(x=m}{P(x=m-}$

∴ $P(x=$

∴ In this case $P(x)$ increase

to decrease. Hence dist. is bimod

Case I. If m is not an integer

$m =$

when x takes the value 'a' (whicl

$x = a$ (greatest integer less t

10.2.4. Moment Generating Fu

By def.,

M_0

Ex. 10-34. *For Poisson distribution show that*

$$m\sigma\gamma_1\gamma_2 = 1.$$

Sol.

$$\text{L.H.S.} = m\sqrt{m}\,\frac{1}{\sqrt{m}}\cdot\frac{1}{m} = 1.$$

10.2.3. Mode of the Poisson Distribution

In Poisson distribution the probability of x successes is given by

$$P(x) = e^{-m}\cdot\frac{m^x}{x!}$$

The mode is that value of x for which $P(x)$ is greater than or equal to $P(x-1)$ and $P(x+1)$ *i.e.*,

$$P(x-1) \le P(x) \ge P(x+1)$$

Consider $$P(x-1) \le P(x)$$

or $$e^{-m}\,\frac{m^{x-1}}{(x-1)!} \le e^{-m}\,\frac{m^x}{x!}$$

or $$x \le m \qquad \ldots(i)$$

Similarly other inequality gives $x \ge m-1$ $\qquad \ldots(ii)$

From (i) and (ii) modal value x satisfies the inequality

$$m-1 \le x \le m \qquad \ldots(iv)$$

Case I. If m is an integer, then $(m-1)$ is also an integer.

Now $$\frac{P(x=m)}{P(x=m-1)} = e^{-m}\,\frac{m^m}{m!}\cdot\frac{(m-1)!}{e^{-m}m^{m-1}}$$

$$= \frac{m}{m} = 1$$

$$\therefore \qquad P(x=m) = P(x=m-1) \qquad \ldots(iv)$$

$\therefore$ In this case $P(x)$ increases till $x = m-1$ and then (iv) holds and after that it begins to decrease. Hence dist. is bimodal with modes $m-1$ and m.

Case I. If m is not an integer, let

$$m = a \text{ (an integer)} + f \text{ (a fraction)}$$

when x takes the value 'a' (which is less than m but greater than $m-1$) from (i) and (ii)

$$P(a-1) < P(a) > P(a+1)$$

$x = a$ (greatest integer less than m) is the mode.

10.2.4. Moment Generating Function

By def.,

$$M_0(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx}\cdot e^{-m}\cdot\frac{m^x}{x!}$$

$$= e^{-m}\sum_{x=0}^{\infty}\frac{(me^t)^x}{x!} = e^{-m}\cdot e^{me^t} = e^{m(e^t-1)}$$

Now $$M_{\bar{x}}(t) = E\{e^{t(x-m)}\} = e^{-mt}E(e^{tx})$$
$$= e^{-mt}.M_0(t) = e^{m(e^t-1-t)}$$

Deduction of Moments

$$M_{\bar{x}}(t) = e^{m(e^t-1-t)}$$
$$= e^{m\left\{\frac{t^2}{2!}+\frac{t^3}{3!}+\dots\dots\right\}}$$
$$= 1+m\left\{\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+\dots\dots\right\}+\frac{m^2}{2!}\left\{\frac{t^2}{2!}+\frac{t^3}{3!}+\dots\right\}^2$$
$$+\frac{m^3}{3!}\left\{\frac{t^2}{2!}+\frac{t^3}{3!}+\dots\right\}^3+\frac{m^4}{4!}\left\{\frac{t^2}{2!}+\dots\dots\right\}^4+\dots\dots$$

But $$M_{\bar{x}}(t) = 1+\mu_1 t+\mu_2\frac{t^2}{2!}+\mu_3\frac{t^3}{3!}+\mu_4\frac{t^4}{4!}+\dots\dots$$

$\therefore$ $$\mu_1 = 0, \mu_2 = m, \mu_3 = m, \mu_4 = m+3m^2.$$

Ex. 10-35. *Show that the sum of two independent Poisson variates is a Poisson variate.*

Sol. Let x_1 and x_2 be two Poisson variates with means m_1 and m_2.

Let $$x = x_1+x_2$$

Then $$M_0(t) \text{ of } x = \{M_0(t) \text{ of } x_1\}.\{M_0(t) \text{ of } x_2\}$$

Now $$M_0(t) \text{ of } x_1 = e^{m_1(e^t-1)}$$

and $$M_0(t) \text{ of } x_2 = e^{m_2(e^t-1)}$$

$\therefore$ $$M_0(t) \text{ of } x = e^{(m_1+m_2)(e^t-1)}$$

which is a m.g.f. of a Poisson variate with mean m_1+m_2.

$\therefore$ x is a Poisson variate with mean m_1+m_2.

Ex. 10-36. *Find $M_0(t)$ of the difference of two independent Poisson variates with means m_1 and m_2 and show that it is not a Poisson variate.*

Sol. Let $$u = x-y$$

Then $$M_0(t) \text{ of } u = E(e^{tu}) = E(e^{t(x-y)})$$
$$= E(e^{tx})E(e^{-ty})$$
$$= e^{m_1(e^t-1)}.e^{m_2(e^{-t}-1)}$$
$$= \exp\{m_1(e^t-1)+m_2(e^{-t}-1)\}$$

Since $M_0(t)$ of u is not of the form $e^{m(e^t-1)}$, u is not a Poisson variate.

10.2.5. Cumulative Function a

By def. Cumulative functio

K

$\therefore$ k

Ex. 10-37. *If x is a Poisson*

P(x

Find mean and variance. A

Sol. Let λ be the parameter

Then

Now P(x

$\therefore$ e

$\therefore$

$\therefore$ Mean = Variance = λ =

Ex. 10-38. *If x and y are in*

P(x

and P(y

find var (

Sol. Let m_1, m_2 be parame

respectively are

$\therefore$ (1) $\Rightarrow$ e^{-m_1}

$\Rightarrow$

and (2) $\Rightarrow$ e^{-m_2}

$\Rightarrow$

$\therefore$ Var

10.2.5. Cumulative Function and Cumulants

By def. Cumulative function is given by

$$K_0(t) = \log M_0(t) = \log e^{m(e^t-1)}$$

$$= m(e^t - 1)$$

$$= m\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots\right)$$

$$\therefore \quad k_1(0) = m = k_2 = k_3 = \ldots\ldots$$

Ex. 10-37. *If x is a Poisson variate such that*

$$P(x=1) = 2P(x=2)$$

Find mean and variance. Also find $P(x=0)$.

Sol. Let λ be the parameter of x.

Then $$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

Now $$P(x=1) = 2P(x=2)$$

$$\therefore \quad e^{-\lambda}.\lambda = 2e^{-\lambda}\frac{\lambda^2}{2!}$$

$$\therefore \quad \lambda = 1$$

$\therefore$ Mean = Variance = $\lambda = 1$ and $P(0) = e^{-1}$.

Ex. 10-38. *If x and y are independent Poisson variates such that*

$$P(x=1) = P(x=2) \qquad \ldots(1)$$

and $$P(y=2) = P(y=3) \qquad \ldots(2)$$

find var $(x-2y)$.

Sol. Let m_1, m_2 be parameters for x, y respectively. Then probability $f^n s$ for x and y respectively are

$$P(x) = e^{-m_1}\frac{m_1^x}{x!}$$

$$P(y) = e^{-m_2}\frac{m_2^y}{y!}$$

$$\therefore (1) \Rightarrow \quad e^{-m_1}.\frac{m_1}{1!} = e^{-m_1}\frac{m_1^2}{2!}$$

$$\Rightarrow \quad m_1 = 2$$

$$\text{and } (2) \Rightarrow \quad e^{-m_2}.\frac{m_2{}^2}{2!} = e^{-m_2}\frac{m_2{}^3}{3!}$$

$$\Rightarrow \quad m_2 = 3.$$

$$\therefore \quad \text{Var }(x) = m_1 = 2$$

and $\quad \text{Var}(y) = m_2 = 3$

let $\quad u = x - 2y$

$\therefore \quad \bar{u} = \bar{x} - 2\bar{y}$

$\therefore \quad u - \bar{u} = (x - \bar{x}) - 2(y - \bar{y})$

$\therefore$
$$\begin{aligned}
\text{Var}(u) &= E(u - \bar{u})^2 \\
&= E\{(x - \bar{x}) - 2(y - \bar{y})\}^2 \\
&= E\{(x - \bar{x})^2 + 4(y - \bar{y})^2 - 4(x - \bar{x})(y - \bar{y})\} \\
&= E(x - \bar{x})^2 + 4E(y - \bar{y})^2 - 4E(x - \bar{x})(y - \bar{y})
\end{aligned}$$

Since x and y are independent,

$$\begin{aligned}
E(x - \bar{x})(y - \bar{y}) &= E\{x \cdot y - \bar{x}y - x\bar{y} + \bar{x} \cdot \bar{y}\} \\
&= E(xy) - \bar{x} \cdot \bar{y} \\
&= E(x)E(y) = \bar{x} \cdot \bar{y} \\
&= \bar{x} \cdot \bar{y} - \bar{x} \cdot \bar{y} = 0
\end{aligned}$$

$\therefore \quad \text{Var}(u) = \text{Var}(x) + 4\ \text{Var}(y).$

$\text{Var}(u) = 14.$

Ex. 10-39. *If x is a Poisson variate with mean m, find*

(*i*) $E(e^{-kx})$.

(*ii*) $E(xe^{-kx})$

Sol. (*i*)
$$\begin{aligned}
E(e^{-kx}) &= \sum_{x=0}^{\infty} e^{-m} \frac{m^x}{x!} e^{-kx} \\
&= e^{-m} \sum_{x=0}^{\infty} \frac{(me^{-k})^x}{x!} = e^{-m} . e^{me^{-k}} \\
&= e^{-m(1-e^{-k})}.
\end{aligned}$$

(*ii*)
$$\begin{aligned}
E(xe^{-kx}) &= \sum_{x=0}^{\infty} e^{-m} \frac{m^x}{x!} xe^{-kx} \\
&= e^{-m} \sum_{x=1}^{\infty} \frac{m^x}{(x-1)!} e^{-kx} \\
&= e^{-m} \sum_{x=1}^{\infty} \frac{(me^{-k})^x}{(x-1)!} \\
&= e^{-m} . (me^{-k}) e^{me^{-k}} \\
&= me^{-m(1-e^{-k})-k}
\end{aligned}$$

Ex. 10-40. *Show that in a*

mean is $\frac{2}{e}$.

Sol. Since in Poisson distri

$\therefore$

$\therefore$ Poiss

$\therefore$ Mean deviation about m

Ex. 10-41. *If x and y are*

Prove that the probability th

exp. $\{mt + m't^{-1} - m - m'\}$.

Sol. $P(x - y = r)$ is require

Now $x - y$ will take the va

where $s = 0, 1, 2, \ldots$

By compound prob. theorem

is

$\therefore$ By total prob. theorem pr

Ex. 10-40. *Show that in a Poisson distribution with unit mean, mean deviation about mean is* $\frac{2}{e}$.

Sol. Since in Poisson distribution with parameter m, mean is m.

$\therefore$ $m = 1$

$\therefore$ Poisson distribution is

$$P(x) = \frac{e^{-1}}{x!}, x = 0, 1, 2 \ldots$$

$\therefore$ Mean deviation about mean is given by

$$\begin{aligned}
\text{M.D.} &= E|x-1| \\
&= \sum_{x=0}^{\infty} |x-1| \frac{e^{-1}}{x!} \\
&= e^{-1} + \sum_{x=2}^{\infty} (x-1) \frac{e^{-1}}{x!} \\
&= e^{-1} + e^{-1} \sum_{x=2}^{\infty} \left\{ \frac{1}{(x-1)!} - \frac{1}{x!} \right\} \\
&= e^{-1} \left\{ 1 + \frac{1}{1!} \right\} \\
&= \frac{2}{e}.
\end{aligned}$$

Ex. 10-41. *If x and y are poisson variates with means m and m' respectively. Prove that the probability that $(x - y)$ has the value r is the co-efficient of t^r in exp. $\{mt + m't^{-1} - m - m'\}$.*

Sol. $P(x - y = r)$ is required.

Now $x - y$ will take the value r when x takes the value $r + s$ and y takes the value s where $s = 0, 1, 2, \ldots$

By compound prob. theorem, prob. of x taking the value $r + s$ and y taking the value s is

$$\left(e^{-m} \frac{m^{r+s}}{(r+s)!} \right) \left(e^{-m'} \frac{m'^s}{s!} \right)$$

$\therefore$ By total prob. theorem prob. of $(x - y)$ taking the value r.

$$= e^{-(m+m')} \sum_{s=0}^{\infty} \frac{m^{r+s} m'^s}{s!(r+s)!}$$

$$e^{-(m+m')}, \text{co-efficient of } t^r \text{ in } e^{mt+m't^{-1}}$$

$$= \text{co-efficient of } t^r \text{ in } e^{mt+m't^{-1}-m-m'}$$

Ex. 10-42. *If x is a Poisson variate with mean m, find M.G.F. of* $z = \dfrac{x-m}{\sqrt{m}}$ *and find its limit when* $m \to \infty$.

Sol.

$$M_0(t) \text{ of } z = E(e^{tz})$$

$$= E\left\{e^{t\left\{\frac{x-m}{\sqrt{m}}\right\}}\right\} = e^{-t\sqrt{m}}\,E\left\{e^{\frac{tx}{\sqrt{m}}}\right\}$$

$$= e^{-t\sqrt{m}}\,M_0\left(\frac{t}{\sqrt{m}}\right) \text{ of } x$$

$$= e^{-t\sqrt{m}}\,e^{m(e^{\frac{t}{\sqrt{m}}}-1)}$$

$$= e^{m(e^{\frac{t}{\sqrt{m}}}-1)-t\sqrt{m}}$$

$\therefore$

$$\log\{M_0(t) \text{ of } z\} = m(e^{\frac{t}{\sqrt{m}}}-1)-t\sqrt{m}$$

$$= m\left\{\frac{t}{\sqrt{m}}+\frac{1}{2!}\left(\frac{t}{\sqrt{m}}\right)^2+\frac{1}{3!}\left(\frac{t}{\sqrt{m}}\right)^3+\ldots\ldots\right\}-t\sqrt{m}$$

$$= \frac{1}{2}t^2 + \text{terms containing } \frac{t}{\sqrt{m}} \text{ and higher powers}$$

$\therefore$

$$\underset{m\to\infty}{\text{Lt}} \log\{M_0(t) \text{ of } z\} = \frac{1}{2}t^2$$

$\therefore$

$$\underset{m\to\infty}{\text{Lt}} M_0(t) \text{ of } z = e^{\frac{1}{2}t^2}.$$

Ex. 10-43. *If x is a P.V. with parameter m show that*

$$P(x>r) < \frac{m^r}{r!}, r = 0, 1, 2, \ldots$$

Sol. We have

$$P(x>r) = \sum_{x=r+1}^{\infty} P(x)$$

$$= \sum_{x=r+1}^{\infty} e^{-m}\frac{m^x}{x!}$$

$$= e^{-m}\left\{\frac{m^{r+1}}{(r+1)!}+\frac{m^{r+2}}{(r+2)!}+\ldots\ldots\right\}\ldots \qquad \ldots(1)$$

we have $(r +$

$\Rightarrow$

$\therefore$ (1) $\Rightarrow$

Now e'

$\Rightarrow$

$\therefore$ (2) $\Rightarrow$

Ex. 10-44. *The probability that is 0·01. By applying Poisson's appr of 100 items selected at random fron item is* $\frac{2}{e}$.

Sol. Here

$\therefore$

$\therefore$ Prob. of x defective items is

$P(x)$

$\therefore$ $P(x \nmid 1)$

Ex. 10-45. *Find the probability 200 fuses if experience show that 2%*

Sol. Let the presence of a defect

Then p = prob. of success = 0·0

Here $n = 200$.

we have $$(r+i)! = (r+i)(r+i-1)\ldots(r+1)r!$$

$$\geq i(i-1)\ldots1,\, r! = i!.r!$$

$$\Rightarrow \qquad \frac{1}{(r+i)!} \leq \frac{1}{(i!.r!)}$$

$$\therefore \ (1) \Rightarrow \qquad P(x > r) < e^{-m}\left\{\frac{m^{r+1}}{1!.r!} + \frac{m^{r+2}}{2!.r!}\ldots\ldots\right\}$$

$$= e^{-m}\frac{m^r}{r!}\left\{m + \frac{m^2}{2!} + \ldots\ldots\right\} \qquad \ldots(2)$$

Now $$e^m = 1 + m + \frac{m^2}{2!} + \ldots\ldots$$

$$> m + \frac{m^2}{2!} + \ldots\ldots$$

$$\Rightarrow \qquad e^{-m}\left(m + \frac{m^2}{2!} + \ldots\ldots\right) < 1$$

$$\therefore \ (2) \Rightarrow \qquad P(x > r) < \frac{m^r}{r!}.$$

Ex. 10-44. *The probability that an item produced by a certain machine will be defective is* $0{\cdot}01$. *By applying Poisson's approximations show that the probability that random sample of 100 items selected at random from the total output will contain no more than one defective item is* $\frac{2}{e}$.

Sol. Here $\qquad n = 100,\ p = 0{\cdot}01$

$\therefore \qquad m = np = 1$

$\therefore$ Prob. of x defective items is given by

$$P(x) = e^{-m}\frac{m^x}{x!} = \frac{e^{-1}}{x!}$$

$$\therefore \qquad P(x \not> 1) = P(x \leq 1)$$

$$= P(0) + P(1)$$

$$= e^{-1} + e^{-1}$$

$$= \frac{2}{e}.$$

Ex. 10-45. *Find the probability that at most 5 defective fuses will be found in a box of 200 fuses if experience show that 2% of such fuses are defective.*

Sol. Let the presence of a defective fuse in the box be called success.

Then p = prob. of success $= 0{\cdot}02$

Here $n = 200$.

Since n is large and p is small the distribution can be taken to be Poissonian.

$\therefore$ $$m = np = (200)(0 \cdot 02) = 4$$

$\therefore$ $$e^{-m} = e^{-4} = 0 \cdot 0183$$

$\therefore$ $$P(x) = e^{-4} \cdot \frac{4^x}{x!} = (0 \cdot 0183)\frac{4^x}{x!}$$

Required Prob. $$= P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$$

$$= (0 \cdot 0183)\left\{1 + 4 + 8 + \frac{32}{3} + \frac{32}{3} + \frac{128}{15}\right\} = 0 \cdot 78.$$

Ex. 10-46. *In a certain factory turning out razor blades, there is a small chance* $\frac{1}{500}$ *for any blade to be defective. The blades are supplied in packets of 10. Use Poisson's distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.*

Sol. Here $$p = \frac{1}{500}, n = 10, N = 10{,}000$$

$\therefore$ $$m = np = \frac{1}{500} \cdot 10 = \frac{1}{50} = 0 \cdot 02$$

$$e^{-m} = e^{-0.02} = 0 \cdot 9802$$

$\therefore$ Prob. of having x defective blades is given by

$$P(x) = \frac{(0 \cdot 9802)(0 \cdot 02)^x}{x!}$$

$\therefore$ No. of packets containing x defective blades

$$= 10{,}000 \frac{(0 \cdot 9802)(0 \cdot 02)^x}{x!}$$

$\therefore$ No. of packets containing no. defective blades

$$= 10{,}000\,(0 \cdot 9802) = 9802$$

No. of packets containing one defective blades

$$= 10{,}000\,(0 \cdot 9802)\,(0 \cdot 02) = 196.04$$

$$\approx 196$$

and No. of packets containing two defective blades

$$= 10{,}000\,(0 \cdot 9802)\,\frac{(0 \cdot 02)^2}{2!} = 1 \cdot 9604$$

$$\approx 2.$$

Ex. 10-47. *Fit Poisson's distribution to the following and calculate theoretical frequencies :*

Death	0	1	2	3	4
Frequencies	122	60	15	2	1

Sol. $$m = \text{mean} = \frac{(122)0 + (60)1 + (15)2 + (2)3 + (1)4}{200}$$

$\therefore$ $e^{-m} = e^{-0.5} = 1 + (-0 \cdot 5) + \frac{1}{2!}$(–

$= 1 - 0 \cdot 5 + 0 \cdot 125 - 0 \cdot 0208 +$(

$= 0 \cdot 61$ (nearly)

$\therefore$ Theoretical frequency of x

$\therefore$ Theoretical frequencies are

122, 61, 15, 2 a

Ex. 10-48. *A car-hire firm has demands for a car on each day is Calculate the proportion of days on which some demand is refused* (e^{-1}

Sol. Let x be the number of de

Then dist of x is

$P(x)$

Now the proportion of days on

and the proportion of days on v

Ex. 10-49. *For a Poisson varia*

$\lambda\{^r c_1 \mu_{r-1}$

Sol. By def.

μ_{r+1}

ken to be Poissonian.

4

4^x

$x!$

- $P(3) + P(4) + P(5)$

$+\frac{32}{3}+\frac{32}{3}+\frac{128}{15}\Big\} = 0\cdot 78.$

's, there is a small chance $\frac{1}{500}$

ı packets of 10. Use Poisson's
:ts containing no defective, one
ıment of 10,000 packets.

),000

$= 0\cdot 02$

$)\cdot 02)^x$

9802

$)\cdot 02) = 196.04$

$0\cdot 02)^2$ / $2!$ $= 1\cdot 9604$

wing and calculate theoretical

3 4
2 1

$.5)2+(2)3+(1)4$

0

$$= \frac{60+30+6+4}{200} = 0\cdot 5$$

$$\therefore\ e^{-m} = e^{-0\cdot 5} = 1+(-0\cdot 5)+\frac{1}{2!}(-0\cdot 5)^2+\frac{1}{3!}(-0\cdot 5)^3+\frac{1}{4!}(-0\cdot 5)^4+\frac{1}{5!}(-0\cdot 5)^5+\ldots\ldots\ldots$$

$$= 1-0\cdot 5+0\cdot 125-0\cdot 0208+0\cdot 0026-0\cdot 00026$$

$$= 0\cdot 61 \text{ (nearly)}$$

$\therefore$ Theoretical frequency of x deaths is

$$200.e^{-0\cdot 5}\frac{(0\cdot 5)^x}{x!}$$

$$= 200.(0\cdot 61)\frac{(0\cdot 5)^x}{x!}$$

$\therefore$ Theoretical frequencies are

122, 61, 15, 2 and 0.

Ex. 10-48. *A car-hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1·5. Calculate the proportion of days on which neither car is used and the proportion of days on which some demand is refused* ($e^{-1\cdot 5} = 0\cdot 2231$).

Sol. Let x be the number of demands for a car in a day.

Then dist of x is

$$P(x) = e^{-1\cdot 5}\frac{(1\cdot 5)^x}{x!}$$

Now the proportion of days on which neither car is used

$$= P\ \{\text{of no. demand in a day}\}$$

$$= P\ (x=0) = e^{-1\cdot 5} = 0\cdot 2231$$

and the proportion of days on which some demand is refused

$$= P\,(x>2)$$

$$= 1-P\,\{x\le 2\}$$

$$= 1-P(x=0)-P(x=1)-P\{x=2)\}$$

$$= 1-e^{-1\cdot 5}\{1+1\cdot 5+1\cdot 125\}$$

$$= 1-(0\cdot 2231)\,\{3\cdot 625\} = 0\cdot 19126.$$

Ex. 10-49. *For a Poisson variate with parameter* λ, *show that*

$$\lambda\{{}^rc_1\,\mu_{r-1}+{}^rc_2\,\mu_{r-2}\ldots+{}^rc_r\mu_0\} = \mu_{r+1}.$$

Sol. By def.

$$\mu_{r+1} = E(x-\lambda)^{r+1}$$

$$= \sum_{x=0}^{\infty}(x-\lambda)^{r+1}e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty}(x-\lambda)^r e^{-\lambda}\frac{\lambda^x}{x!}(x-\lambda)$$

$$= \sum_{x=0}^{\infty} x(x-\lambda)^r e^{-\lambda}\frac{\lambda^x}{x!} - \lambda\sum_{x=0}^{\infty}(x-\lambda)^r e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= \sum_{x=1}^{\infty}(x-1-\lambda+1)^r e^{-\lambda}\frac{\lambda^x}{(x-1)!} - \lambda\mu_r$$

$$= \lambda\sum_{x=0}^{\infty}(x-\lambda+1)^r e^{-\lambda}\frac{\lambda^x}{x!} - \lambda\mu_r \quad \text{(changing } x \text{ to } x+1\text{)}$$

$$= \lambda\sum_{x=0}^{\infty}\{(x-\lambda)^r + {}^r c_1 (x-\lambda)^{r-1} - \ldots + {}^r c_r\} e^{-\lambda}\frac{\lambda^x}{x!} - \lambda\mu_r$$

$$= \lambda[\mu_r + {}^r c_1 \mu_{r-1} + \ldots\ldots {}^r c_r] - \lambda\mu_r$$

$$= \lambda[{}^r c_1 \mu_{r-1} + \ldots + {}^r c_r \mu_0]$$

$$(\because \mu_0 = 1)$$

Ex. 10-50. *If x and y are independent Poisson variates, show that the conditional distribution of x given x + y is binomial.*

Sol. Let λ, μ be the parameters of x, y respectively.

Then, $z = x + y$ is a P.V. with parameter $\lambda + \mu$. Distributions of x, y, z are

$$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

$$P(y) = e^{-\mu}\frac{\mu^y}{y!}$$

$$P(z) = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^z}{z!}$$

Now $P(x = r / z = n) = \dfrac{P(x = r, z = n)}{P(z = n)}$

$$= \frac{P(x = r, x + y = n)}{P(z = n)}$$

$$= \frac{P(x = r, y = n - r)}{P(z = n)}$$

$$= \frac{P(x = r)\, P(y = n - r)}{P(z = n)}$$

$\{\because x, y$ are independent$\}$

$$= \frac{e^{-\lambda}.\dfrac{\lambda^r}{r!}.e^{-\mu}.\dfrac{\mu^{n-r}}{(n-r)!}}{e^{-(\lambda+\mu)}\dfrac{(\lambda+\mu)^n}{n!}}$$

$= \frac{\ }{r!}$

$= {}^n C$

where $p = \frac{\ }{\lambda}$

which gives the conditional dis

n and $p = \dfrac{\lambda}{\lambda+\mu}$.

Ex. 10-51. *A telephone s hour. The board can make a distribution to find the probab*

Sol. Let x donate the no. o

m = ave

$= \frac{60}{60}$

$\therefore \quad P(x) = e^{-1}$

$\therefore$ Reqd. prob. $= P(x$

$= 1-$

1. For a Poisson distribution

2. If x is the number of occu

$P(x$

3. If x is a P.V. s.t.

$P(x$

find (*i*) m, mean of x(*ii*) β

4. Examine, if the following distribution :

Mean = 1·5 cm, Variance =

$\mu_3 = 1\cdot5\ \text{cm}^3$, $\mu_4 = 8\cdot25$

5. In a certain factory turning

lens to be defective. The le to calculate the approximat two defective, three defecti

$e^{-0.02} = 0\cdot9802$) (See 10-

$-\lambda)^r e^{-\lambda} \frac{\lambda^x}{x!}$

$-\lambda\mu_r$

(changing x to $x+1$)

$-...+^r c_r\} e^{-\lambda} \frac{\lambda^x}{x!} - \lambda\mu_r$

μ_r

$(\because \mu_0 = 1)$

riates, show that the conditional

tributions of x, y, z are

$\{\because x, y$ are independent$\}$

$$= \frac{n!}{r!(n-r)!}\left(\frac{\lambda}{\lambda+\mu}\right)^r \left(\frac{\mu}{\lambda+\mu}\right)^{n-r}$$

$$= {}^n c_r \, p^r \, q^{n-r}$$

where $$p = \frac{\lambda}{\lambda+\mu}, q = \frac{\mu}{\lambda+\mu}$$

which gives the conditional distribution of x given $x+y=n$. This is a B.D. with parameters n and $p = \frac{\lambda}{\lambda+\mu}$.

Ex. 10-51. *A telephone switchboard handles 600 calls on an average during a rush hour. The board can make a maximum of 20 connections per minute. Use the Poisson distribution to find the probability that the board will be overtaxed during any given minute.*

Sol. Let x donate the no. of calls per minute. x is a P.V. with mean m, where

m = average no. of calls which board can handle in a minute

$$= \frac{600}{60} = 10.$$

$$\therefore \quad P(x) = e^{-10} . \frac{(10)^x}{x!}$$

$\therefore$ Reqd. prob. $= P(x > 20) = 1 - P(x \le 20)$.

$$= 1 - \sum_{x=0}^{20} e^{-10} . \frac{(10)^x}{x!}.$$

EXERCISES

1. For a Poisson distribution with parameter λ show that

$$P(x+1) = \frac{\lambda}{x+1} P(x).$$

2. If x is the number of occurrences of the Poisson variate with mean m, show that

$$P(x=n) - P(x=n+1) = P(x=n)$$

3. If x is a P.V. s.t.

$$P(x=2) = 9P(x=4) + 90P(x=6)$$

find (*i*) m, mean of x (*ii*) β_1 the co-efficient of skewness.

4. Examine, if the following are consistent to be the first four moments of a Poisson distribution :
Mean = 1·5 cm, Variance = 1·5 cm^2
$\mu_3 = 1 \cdot 5 \text{ cm}^3, \mu_4 = 8 \cdot 25 \text{ cm}^4$.

5. In a certain factory turning out optical lenses there is a small chance $\frac{1}{500}$ for any one lens to be defective. The lenses are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective, one defective, two defective, three defective lenses in a consignment of 20,000 packets. (Given that $e^{-0.02} = 0 \cdot 9802$) (See 10-46).

6. Find the mean and standard deviation for the table of deaths of women over 85-year old recorded in a three years period :

No. of deaths recorded in a day	0	1	2	3	4	5	6	7
No. of days	364	376	218	89	33	13	2	1

Find the expected number of days with one death recorded from the Poisson series fitted to the data. **[Ans.** 1·18, 1·17, 397]

7. Red blood cell deficiency may be determined by examining a specimen of the blood under a microscope. Suppose a certain small fixed volume contains on the average 20 red cells for normal persons. Using Poisson distribution, obtain the probability that a specimen from a normal person will contain less than 15 red cells.

$$\left[\textbf{Ans. } e^{-20}\sum_{x=0}^{14}\frac{(20)^x}{x!}\right]$$

8. A large number of observations on a given solution, which contained bacteria, were made taking samples of 1 c.c. each and noting down the number of bacteria present in each sample. Assuming the Poisson distribution and given that 10% samples contained no bacteria, find the average number of bacteria per c.c. **[Ans.** 2·3026]

9. In 1000 extensive sets of trials for an event of small probability the frequencies 'f' of the number x of successes are found to be

x:	0	1	2	3	4	5	6	7
f:	305	365	210	80	28	9	2	1

Fitting Poisson distribution to the above data, calculate theoretical frequencies. [**Ans.** 301, 361, 217, 87, 26, 6, 1 and 2]

10. The following data gives the frequency distribution of the number of men killed by the kick of a horse in 10 Prussian Army Corps per army corps per annum over 20 years.

No. of deaths	0	1	2	3	4 and over	Total
Frequency	109	65	22	3	1	200

Show that the distribution is roughly Poissonian and calculate the theoretical frequencies. $(e^{-0.61} = 0\cdot5434)$ [**Ans.** 109, 66, 20, 4 and 1 (4 and over)]

11. A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 10 pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality ?

$(e^{-5} = 0\cdot006738)$

$$\left[\textbf{Ans. } 1-(0\cdot0067)\sum_{x=0}^{10}\frac{5^x}{x!}\right]$$

12. Letters were received in an office on each of 100 days. Assuming the following data to form a random sample from a Poisson distribution, find the expected frequencies, correct to the nearest unit. $(e^{-4} = 0\cdot0183)$

No. of letters	0	1	2	3	4	5	6	7	8	9	10
Frequency	1	4	15	22	21	20	8	6	2	0	1

[**Ans.** 2, 7, 15, 20, 20, 16, 10, 6, 3, 1, 1]

13. Six coins are tossed 6400 ti
probability of getting six h

14. An area of 144 square kilo
appeared constant. To test

divided into 576 squares o

squares containing 0, 1, 2,
given below :
No. of flying bombs
per square 0
Actual no. of
square 229
Calculate the theoretical p

10-3. Normal Distribution

Normal Probability Distri
m and s.d. σ *is*

$dP = -$
σ

The variate x is called norm
The curve with equation

is called normal curve.

If N is the total freq. the cor

Def : Standard normal distri

where

(*i*) *To derive normal distribu*
Normal distribution can be
when n, the number of trials is v

eaths of women over 85-year

4	5	6	7
33	13	2	1

rded from the Poisson series
[**Ans.** 1·18, 1·17, 397]
ning a specimen of the blood
ne contains on the average 20
, obtain the probability that a
5 red cells.

$$\left[\textbf{Ans. } e^{-20}\sum_{x=0}^{14}\frac{(20)^x}{x!}\right]$$

hich contained bacteria, were
number of bacteria present in
n that 10% samples contained
[**Ans.** 2·3026]
bability the frequencies '*f*' of

5	6	7
9	2	1

theoretical frequencies.
, 361, 217, 87, 26, 6, 1 and 2]
e number of men killed by the
rps per annum over 20 years.

4 and over	Total
1	200

nd calculate the theoretical
, 66, 20, 4 and 1 (4 and over)]
oduct is defective. If he sells
than 10 pins will be defective,
meet the guaranteed quality ?

$$\left[\textbf{Ans. } 1-(0\cdot 0067)\sum_{x=0}^{10}\frac{5^x}{x!}\right]$$

ssuming the following data to
ind the expected frequencies,

6	7	8	9	10
8	6	2	0	1

, 15, 20, 20, 16, 10, 6, 3, 1, 1]

13. Six coins are tossed 6400 times. Using the Poisson distribution, what is the approximate probability of getting six heads x times.

$$\left[\textbf{Ans. } e^{-100}\frac{(100)^x}{x!}\right]$$

14. An area of 144 square kilometres was selected for which the mean density of bombs appeared constant. To test the hypothesis that the bombs fell in clusters, the area was divided into 576 squares of $\frac{1}{4}$ kilometre each and a count made of the numbers of squares containing 0, 1, 2, etc., bombs of which there were 576 altogether. The data is given below :

No. of flying bombs per square	0	1	2	3	4	5 and over
Actual no. of square	229	211	93	35	7	1

Calculate the theoretical poisson frequencies.

[**Ans.** 227, 211, 98, 31, 7 and 2]

10-3. Normal Distribution

Normal Probability Distribution. *The normal distribution of the variate x with mean m and s.d. σ is*

$$dP = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}dx \quad -\infty < x < \infty.$$

The variate x is called normal variate.
The curve with equation

$$y = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

is called normal curve.

If N is the total freq. the corresponding normal curve is

$$y = \frac{N}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Def : Standard normal distribution function is defined by

$$\Phi(x) = \int_{-\infty}^{x} f(x)dx.$$

where
$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

(*i*) *To derive normal distribution as a limiting form of binomial distribution.*

Normal distribution can be regarded as the limiting form of the binomial distribution when n, the number of trials is very large and neither p nor q is very small.

Let $$z = \frac{x-np}{\sqrt{npq}} \qquad ...(1)$$

where x is a binomial variate with parameters n and p. Since mean and *s.d.* of x are np and $\sqrt{npq}$, the variate z defined by (1) has zero mean and unit variance. As x takes values from 0 to n, z takes values from $-\sqrt{\frac{np}{q}}$ to $\sqrt{\frac{nq}{p}}$ and the jump in the value of z at each stage is $\frac{1}{\sqrt{npq}}$. Now as $n \to \infty$ two extreme values of z tend to $-\infty$ and ∞ respectively and the jump at each stage tends to zero. Thus, in the limit we expect the distribution of z to be continuous extending from $-\infty$ to ∞ and having zero mean and unit variance.

In binomial dist. the prob. for the variate x to take value x is given by

$$P(x) = {}^n c_x p^x q^x = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

By Stirling's formula

$$n! \simeq \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}$$

$$\therefore \quad P(x) \simeq \frac{\sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}}{(\sqrt{2\pi} e^{-x}.x^{x+\frac{1}{2}})(\sqrt{2\pi} e^{-n+x}.(n-x)^{n-x+\frac{1}{2}})} p^x q$$

$$\cong \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}$$

Let $$N = \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}$$

$$\therefore \quad \log N = -\left(x+\frac{1}{2}\right)\log\frac{x}{np} - \left(n-x+\frac{1}{2}\right)\log\frac{n-x}{nq}$$

From (1) $x = np + z\sqrt{npq}$

$$\therefore \quad \log N = -\left(np + z\sqrt{npq} + \frac{1}{2}\right)\log\left(1 + z\sqrt{\frac{q}{np}}\right)$$
$$-\left(nq - z\sqrt{npq} + \frac{1}{2}\right)\log\left(1 - z\sqrt{\frac{p}{nq}}\right)$$

As n is very large and tends to infinity both $z\sqrt{\frac{q}{np}}$ and $z\sqrt{\frac{p}{nq}}$ can be taken to be less than unity and hence both the logarithms can be expanded in series.

$\therefore$

$= -\frac{1}{2}.z^2 +$ terms co

$\therefore \log N \to -\frac{1}{2} z^2$ as

i.e., $N \to e^{-\frac{1}{2}z^2}$ as $N \to \propto$

Since $\frac{1}{\sqrt{npq}}$ is the in

its limit by dz.

$\therefore$ If dP denotes the

$z + \frac{1}{2} dz$ we have

This is the required co

(*ii*) *To derive normal*

Normal distribution ca

when its parameter m is la

Let

where x is a Poisson variat

variate z defined by (1) ha

takes values from $-\sqrt{m}$ to

$m \to \infty$ two extreme valu

zero. Thus in the limit we

to ∞ and having zero mea

x is given by

...(1)

mean and *s.d.* of x are np and

iriance. As x takes values from

the value of z at each stage is

∞ and ∞ respectively and the

ect the distribution of z to be

and unit variance.

x is given by

$p^x q^{n-x}$

$$\frac{{}^{-n} n^{n+\frac{1}{2}}}{\pi e^{-n+x} . (n-x)^{n-x+\frac{1}{2}})} p^x q$$

$$\frac{1}{2}\left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}$$

$+\frac{1}{2}$

$$n-x+\frac{1}{2}\bigg) \log \frac{n-x}{nq}$$

$$\log\left(1+z\sqrt{\frac{q}{np}}\right)$$

$$-z\sqrt{npq}+\frac{1}{2}\bigg) \log\left(1-z\sqrt{\frac{p}{nq}}\right)$$

$z\sqrt{\frac{p}{nq}}$ can be taken to be less

n series.

$$\therefore \qquad \log N = -\left(np+z\sqrt{npq}+\frac{1}{2}\right)\left[z\sqrt{\frac{q}{np}}-\frac{1}{2}z^2\frac{q}{np}+\ldots\right]$$

$$+\left(nq-z\sqrt{npq}+\frac{1}{2}\right)\left[z\sqrt{\frac{p}{nq}}+\frac{1}{2}z^2\frac{p}{nq}+\ldots\right]$$

$= -\frac{1}{2}.z^2 +$ terms containing n in the denominator

$\therefore$ $\log N \to -\frac{1}{2}z^2$ as $n \to \infty$

i.e., $N \to e^{-\frac{1}{2}z^2}$ as $N \to \infty$.

Baron Simeon Denis Poisson (21 Jun 1781- 25 Apr 1840) French Mathematician and Physicist

Since $\frac{1}{\sqrt{npq}}$ is the increment in z at each stage and tends to zero as $n \to \infty$ we denote its limit by dz.

$\therefore$ If dP denotes the probability for the variate z to lie in the interval $z-\frac{1}{2}dz$ and $z+\frac{1}{2}dz$ we have

$$dP = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz.$$

This is the required continuous distribution of z and is called Normal distribution.

(*ii*) *To derive normal distribution as limiting form of Poisson distribution.*

Normal distribution can also be regarded as the limiting form of the Poisson distribution when its parameter m is large.

Let $$z = \frac{x-m}{\sqrt{m}} \qquad \ldots(1)$$

where x is a Poisson variate with parameter m. Since mean and s.d. of x are m and $\sqrt{m}$, the variate z defined by (1) has zero mean and unit variance. As x takes values from 0 to ∞, z takes values from $-\sqrt{m}$ to ∞ and the jump in the value of z at each stage is $\frac{1}{\sqrt{m}}$. Now as $m \to \infty$ two extreme values of z tend to $-\infty$ and ∞ and the jump at each stage tends to zero. Thus in the limit we expect the distribution of z to be continuous extending from $-\infty$ to ∞ and having zero mean and unit variance. In Poisson dist. the prob. for x to take value x is given by

$$P(x) = e^{-m}\frac{m^x}{x!} \cong e^{-m}\frac{m^x}{\sqrt{2\pi}e^{-x}x^{x+\frac{1}{2}}}$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{m}} e^{x-m} \cdot \left(\frac{m}{x}\right)^{x+\frac{1}{2}}$$

Let $$N = e^{x-m} \cdot \left(\frac{m}{x}\right)^{x+\frac{1}{2}}$$

$\therefore$ $$\log N = x - m - \left(x + \frac{1}{2}\right) \log \frac{x}{m}$$

$$= z\sqrt{m} - \left(m + z\sqrt{m} + \frac{1}{2}\right) \log\left(1 + \frac{z}{\sqrt{m}}\right) \quad \text{[from (1)]}$$

$$= z\sqrt{m} - \left(m + z\sqrt{m} + \frac{1}{2}\right)\left(\frac{z}{\sqrt{m}} - \frac{1}{2}\frac{z^2}{m} + \ldots\right)$$

$$= -\frac{1}{2}z^2 + \text{ terms containing } m \text{ in the denominator}$$

$\therefore$ $\log N \to -\frac{1}{2}z^2$ or $N \to e^{-\frac{1}{2}z^2}$ as $m \to \infty$

Since $\frac{1}{\sqrt{m}}$ is the increment in z at each stage and tends to zero as $m \to \infty$ we denote its limit by dz.

$\therefore$ If dP denotes the probability for the variate z to lie in the interval $z - \frac{1}{2}dz$ and $z + \frac{1}{2}dz$ we have

$$dP = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

10.3.1. Mean deviation about mean for a normal variate with mean m and s.d. σ.

Sol. Dist. of a normal variate x is

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx, -\infty < x < \infty$$

$\therefore$ Mean deviation from mean

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-m| e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |\sigma y| e^{-\frac{1}{2}y^2} dy$$

where $$y = \frac{x-m}{\sigma}$$

10.3.2. Moments

(i) Odd order moments
For a normal variate wi

$\therefore$

(ii) Even order moment
By def.

$\therefore$

Put $n = 2$

$\therefore$ $\mu_4 = 3\sigma^2 \mu_2 = 3\sigma^4.$

$$= \frac{2\sigma}{\sqrt{2\pi}}\int_0^\infty ye^{-\frac{1}{2}y^2}\,dy = \frac{2\sigma}{\sqrt{2\pi}}\left\{-e^{-\frac{1}{2}y^2}\right\}_0^\infty$$

$$= \sigma\sqrt{\frac{2}{\pi}} \cong \frac{4\sigma}{5}.$$

$$= 80\%\ \sigma$$

10.3.2. Moments

(*i*) *Odd order moments about mean*

For a normal variate with mean m and s.d. σ

$$dP = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}dx,\ -\infty < x < \infty$$

$$\therefore \qquad \mu_{2n+1} = E\,(x-m)^{2n+1}$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-m)^{2n+1}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}dx$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y^{2n+1}e^{-\frac{1}{2}y^2}\,dy \qquad \text{where } y = \frac{x-m}{\sigma}$$

$$= 0 \qquad \text{(as integrand is an odd } f^n\text{).}$$

(*ii*) *Even order moments about mean*

By def.

$$\therefore \qquad \mu_{2n} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-m)^{2n}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}dx$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y^{2n}e^{-\frac{1}{2}y^2}\,dy \qquad \text{where } y = \frac{x-m}{\sigma}$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}}\left\{\left|-e^{-\frac{1}{2}y^2}.y^{2n-1}\right|_{-\infty}^{\infty} + (2n-1)\int_{-\infty}^{\infty}y^{2n-2}e^{-\frac{1}{2}y^2}\,dy\right\}$$

$$= \sigma^2(2n-1).\frac{1}{\sqrt{2\pi}}\sigma^{2n-2}\int_{-\infty}^{\infty}e^{-\frac{1}{2}y^2}.y^{2n-2}\,dy$$

$$= \sigma^2(2n-1)\mu_{2n-2}$$

Put $n = 2$

$\therefore\ \mu_4 = 3\sigma^2\mu_2 = 3\sigma^4.$

Ex. 10-52. *Show that*

$$\mu_{2n} = 1.3.5.......(2n-1)\,\sigma^{2n}.$$

Sol. By recurrence formula

$$\mu_{2n} = (2n-1)\sigma^2\,\mu_{2n-2}$$

Put $\quad n = n,\ n-1, ..., 2, 1$

$$\mu_{2n} = (2n-1)\,\sigma^2\,\mu_{2n-2}$$

$$\mu_{2n-2} = (2n-3)\,\sigma^2\,\mu_{2n-4}$$

....

$$\mu_4 = 3\sigma^2\,\mu_2$$

$$\mu_2 = 1.\,\sigma^2\,\mu_0 = 1.\,\sigma^2 \qquad (\because \mu_0 = 1)$$

Multiplying

$$\mu_{2n} = 1.3.......(2n-1)\sigma^{2n}.$$

Ex. 10.53. *Let x be a N(m, σ) {i.e., a normal variate with mean m and s.d. σ}, then.*

(*i*) $$\mu'_{r+2} = 2m\mu'_{r+1} + (\sigma^2 - m^2)\,\mu'_r + \sigma^3\frac{d\mu'_r}{d\sigma}$$

where μ'_r *denotes the rth moment about zero.*

(*ii*) $$\mu_{2r+2} = \sigma^2\mu_{2r} + \sigma^3\frac{d\mu_{2r}}{d\sigma}.$$

Sol. By def.

$$\mu'_r = E(x^r)$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^r e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$\therefore \quad \frac{d\mu'_r}{d\sigma} = -\frac{1}{\sigma^2\sqrt{2\pi}}\int_{-\infty}^{\infty} x^r e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$+\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^r e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.\left\{\frac{(x-m)^2}{\sigma^3}\right\} dx$$

$$= -\frac{\mu'_r}{\sigma} + \frac{1}{\sigma^4\sqrt{2\pi}}\int_{-\infty}^{\infty} x^r(x^2 - 2xm + m^2)\,e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$= \mu'_r\left(\frac{m^2}{\sigma^3} - \frac{1}{\sigma}\right) + \frac{\mu'_{r+2}}{\sigma^3} - \frac{2m}{\sigma^3}\mu'_{r+1}$$

∴ μ

(*ii*) By def.

∴ μ_2

10.3.3. Measures of Skewness a

We have

Since $\gamma_2 = 0$ dist. is called

10.3.4. Moment Generating Fu

By Def.,

M_0

where

$$\therefore \qquad \mu'_{r+2} = 2m\mu'_{r+1} + (\sigma^2 - m^2)\mu'_r + \sigma^3 \frac{d\mu'_r}{d\sigma}$$

(*ii*) By def.

$$\mu_{2r} = E(x-m)^{2r}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^{2r} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$\frac{du_{2r}}{d\sigma} = -\frac{1}{\sigma^2\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^{2r} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$+ \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^{2r} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} . \left\{\frac{(x-m)^2}{\sigma^3}\right\} dx$$

$$= -\frac{\mu_{2r}}{\sigma} + \frac{\mu_{2r+2}}{\sigma^3}$$

$$\therefore \qquad \mu_{2r+2} = \sigma^2 \mu_{2r} + \sigma^3 \frac{du_{2r}}{d\sigma}.$$

10.3.3. Measures of Skewness and Kurtosis

We have

$$\mu_3 = 0, \mu_4 = 3\sigma^4$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2} = 3$$

$$\gamma_1 = \sqrt{\beta_1} = 0 \text{ and } \gamma_2 = \beta_2 - 3 = 0$$

Since $\gamma_2 = 0$ dist. is called **Normal dist.**

10.3.4. Moment Generating Function

By Def.,

$$M_0(t) = E\{e^{tx}\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} . e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(m+\sigma y)} e^{-\frac{1}{2}y^2} dy$$

where $\qquad y = \dfrac{x-m}{\sigma}$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tm} . e^{-\frac{1}{2}(y-t\sigma)^2 + \frac{t^2\sigma^2}{2}} dy$$

$$= e^{tm+\frac{1}{2}t^2\sigma^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

where $$z = y - t\sigma$$

$$= e^{tm+\frac{1}{2}t^2\sigma^2}$$

$\therefore$ M.G.F. about mean m is given by

$$M_{\bar{x}}(t) = E\{e^{t(x-m)}\} = e^{-mt}.E\{e^{tx}\}$$

$$= e^{-mt}.e^{mt+\frac{1}{2}t^2\sigma^2} = e^{\frac{1}{2}t^2\sigma^2}$$

Deduction

$$M_{\bar{x}}(t) = e^{\frac{1}{2}t^2\sigma^2}$$

$$= 1 + \left(\frac{1}{2}t^2\sigma^2\right) + \frac{\left(\frac{1}{2}t^2\sigma^2\right)^2}{2!} + \ldots + \frac{\left(\frac{1}{2}t^2\sigma^2\right)^n}{n!} + \ldots$$

$\therefore$ $$\mu_{2n+1} = 0$$

and $$\frac{\mu_{2n}}{(2n)!} = \frac{1}{2^n}.\frac{\sigma^{2n}}{n!}$$

$\Rightarrow$ $$\mu_{2n} = \frac{1}{2^n}.\frac{(2n)!}{n!}\sigma^{2n}$$

$$(2n-1)\ldots\ldots\ldots\ldots 3.1\,\sigma^{2n}.$$

10.3.5. Cumulative Function and Cumulants

By def., cumulative f^n is given by

$$K_0(t) = \log M_0(t) = \log e^{mt+\frac{1}{2}t^2\sigma^2} = mt + \frac{1}{2}t^2\sigma^2$$

But $$K_0(t) = k_1 t + k_2 \frac{t^2}{2!} + \ldots\ldots.$$

where $k_1, k_2, \ldots\ldots$ are various cumulants.

$\therefore$ $k_1(0) = m, k_2 = \sigma^2, k_3 = 0, k_4 = 0, \ldots\ldots$

Thus, all cumulants after the second are equal to zero.

Ex. 10-54. *Show that a linear combination of independent normal variates is also a normal variate.*

Sol. Let $x_1, x_2, \ldots\ldots x_n$ be independent, normal variates with means $m_1, m_2, \ldots\ldots m_n$ and $s.ds.\ \sigma_1, \sigma_2, \ldots\ldots \sigma_n$.

Let

where a's are constants.

Now $M_0(t)$ of

which is the m.g.f. of a normal var

$\therefore$ u is a normal variate with n

Ex. 10-55. *If the independent v*

the common mean μ, *with a commo*

normally distributed about the san

Sol. Here $m_1 = m_2 = \ldots\ldots = m_n$

$\therefore u = \frac{1}{n}\sum_{i=0}^{n} x_i = \bar{x}$ is a norma

$\frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

Ex. 10-56. *Show that for the N*

Sol. The density curve for the

Put $x - n$

which is evidently symmetrical abc

Let $\quad u = a_1x_1 + a_2x_2 + \ldots\ldots\ldots\ldots a_nx_n$

where ***a***'s are constants.

Now
$$M_0(t) \text{ of } u = E\{e^{t(a_1x_1+a_2x_2 \ldots\ldots\ldots\ldots +a_nx_n)}\}$$
$$= E(e^{a_1tx_1}).E(e^{a_2tx_2})\ldots\ldots E(e^{a_ntx_n})$$
$$\{\because (x_1, x_2, \ldots. \text{are independent}\}$$
$$= M_0(ta_1) \text{ of } x_1 . M_0(ta_2) \text{ of } x_2 \ldots.. M_0(a_nt) \textbf{ of } \boldsymbol{x_n}$$
$$= e^{\{ta_1m_1+\frac{1}{2}(ta_1)^2\sigma_1{}^2\}} . e^{\{(ta_2)m_2+\frac{1}{2}(ta_2)^2\sigma_2^2\}}$$
$$\ldots\ldots\ldots\ldots e^{\{(ta_n)m_n+\frac{1}{2}(ta_n)^2\sigma_n{}^2\}}$$
$$= e^{t\Sigma a_i m_i+\frac{1}{2}t^2\Sigma a_i^2\sigma_i^2}$$

which is the m.g.f. of a normal variate with **mean** $\boldsymbol{\Sigma a_i m_i}$ and variance $\sum a_i^2\sigma_i^2$.

$\therefore$ u is a normal variate with mean $\Sigma a_i m_i$ and variance $\Sigma a_i^2\sigma_i^2$.

Ex. 10-55. *If the independent variates $x_i (i = 1, 2, \ldots\ldots n)$ are normally distributed about the common mean μ, with a common variance σ^2, show that their mean $\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$ is also normally distributed about the same mean μ but with variance $\frac{\sigma^2}{n}$.*

Sol. Here $m_1 = m_2 = \ldots.. = m_n = \mu$ and $a_1 = a_2 = \ldots.. = a_n = \frac{1}{n}$

$\therefore u = \frac{1}{n}\sum_{i=0}^{n} x_i = \bar{x}$ is a normal variate with mean $\Sigma a_i m_i = \frac{\Sigma\mu}{n} = \mu$ and variance $=$

$\frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

Ex. 10-56. *Show that for the N.D. mean, mode and median coincide.*

Sol. The density curve for the N.D. is
$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Put $\quad x - m = X$
$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{X^2}{\sigma^2}}$$

which is evidently symmetrical about the line $X = 0$ *i.e.*, $x = m$.

$\therefore$ $x = m$ is the median.

Also evidently y decreases continuously as X increases numerically and is maximum for $X = 0$.

$\therefore X = 0$ *i.e.*, $x = m$ is the mode.

$\therefore$ Mean = Mode = Median = m.

Ex. 10-57. *Find the points of inflextion of the normal curve.*

Sol. The eq. of the normal curve is

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \qquad \ldots(1)$$

At the points of inflextion

$$\frac{d^2y}{dx^2} = 0 \text{ and } \frac{d^3y}{dx^3} \neq 0$$

From (1) $$\frac{d^2y}{dx^2} = \frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}\left\{-\frac{x-m}{\sigma^2}\right\}^2 - \frac{1}{\sigma^2} e^{-\frac{1}{2}\left\{\frac{x-m}{\sigma}\right\}^2}\right]$$

and $$\frac{d^3y}{dx^3} = \frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}\left\{-\frac{x-m}{\sigma^2}\right\}^3 + \frac{3(x-m)}{\sigma^4} e^{-\frac{1}{2}\left\{\frac{x-m}{\sigma}\right\}^2}\right]$$

Put $$\frac{d^2y}{dx^2} = 0$$

$\therefore$ $$\frac{(x-m)^2}{\sigma^2} - 1 = 0$$

or $$x = m \pm \sigma.$$

At $$x = m \pm \sigma, \quad \frac{d^3y}{dx^3} = \frac{1}{\alpha\sqrt{2\pi}} e^{-\frac{1}{2}}\left\{\mp\frac{1}{\sigma^3} \pm \frac{3}{\sigma^3}\right\}$$

$$= \pm\frac{2}{\sigma^4\sqrt{2\pi}} e^{-\frac{1}{2}} \neq 0$$

$\therefore$ At the points of inflextion

$$x = m \pm \sigma$$

and hence from (1) $$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}.$$

Ex. 10-58. *Give chief features of the normal curve.*

Sol. The eq. of the normal curve is

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \qquad \ldots(1)$$

(*i*) Mean, mode and median c

(*ii*) Since y becomes zero whe negative and positive sides, at infir and positive sides.

(*iii*) At the points of inflextior

Evidently the points of inflext

(*iv*) Maximum value of ordina

Ex. 10-59. *Deduce the first fo from those of* (*i*) *the Binomial dist.*

Sol. (*i*) For B.D. μ

μ

μ

Let

when x is a binomial variate with p

Then mean of

μ_2 for $z = E(z-0)^2$

μ_3 for $z = E(z^3)$

μ_4 for $z = E(z^4)$

Now as $n \to \infty, z \to a$ normal

$\therefore$ μ_2 for normal variate

$\therefore$ μ_3 for normal variate

$\therefore$ μ_4 for normal variate

(*i*) Mean, mode and median of the normal curve coincide.

(*ii*) Since y becomes zero when x is numerically infinite, curve touches x-axis both on negative and positive sides, at infinity *i.e.*, x-axis is asymptote to the curve both on negative and positive sides.

(*iii*) At the points of inflextion

$$x = m \pm \sigma$$

Evidently the points of inflextion are equidistant from $x = m$.

(*iv*) Maximum value of ordinate is

$$y = \frac{1}{\sigma\sqrt{2\pi}}.$$

Ex. 10-59. *Deduce the first four moments about the mean of the normal distribution from those of* (*i*) *the Binomial dist.* (*ii*) *the Poisson distribution.*

Sol. (*i*) For B.D.

$$\mu_2 = npq$$

$$\mu_3 = npq(q-p)$$

$$\mu_4 = npq\{1+3(n-2)pq\}$$

Let

$$z = \frac{x-np}{\sqrt{npq}}$$

when x is a binomial variate with parameters n and p.

$$\text{Then mean of } z = E(z) = \frac{E(x-np)}{\sqrt{npq}} = 0$$

$$\mu_2 \text{ for } z = E(z-0)^2 = \frac{1}{npq}E(x-np)^2 = 1$$

$$\mu_3 \text{ for } z = E(z^3) = \frac{1}{(npq)^{\frac{3}{2}}}E(x-np)^3 = \frac{q-p}{\sqrt{npq}}$$

$$\mu_4 \text{ for } z = E(z^4) = \frac{1}{(npq)^2}E(x-np)^4 = \frac{1+3(n-2)pq}{npq}$$

Now as $n \to \infty$, $z \to a$ normal variate

$$\therefore \quad \mu_2 \text{ for normal variate} = \underset{n\to\infty}{\text{Lt}}\ \mu_2 \text{ for } z = 1$$

$$\therefore \quad \mu_3 \text{ for normal variate} = \underset{n\to\infty}{\text{Lt}}\ \mu_3 \text{ for } z = \underset{n\to\infty}{\text{Lt}}\ \frac{q-p}{\sqrt{npq}} = 0$$

$$\therefore \quad \mu_4 \text{ for normal variate} = \underset{n\to\infty}{\text{Lt}}\ \mu_4 \text{ for } z$$

$$= \underset{n\to\infty}{\text{Lt}}\left\{\frac{1-6pq}{npq}+3\right\} = 3.$$

(*ii*) For P.D. $\mu_2 = m$

$$\mu_3 = m$$

$$\mu_4 = m + 3m^2$$

Let $$z = \frac{x-m}{\sqrt{m}}$$

where x is a Poisson variate with parameter m.

$$\text{Then mean of } z = \frac{E(x-m)}{\sqrt{m}} = 0$$

$$\mu_2 \text{ for } z = E(z)^2 = \frac{1}{m}E(x-m)^2 = 1$$

$$\mu_3 \text{ for } z = E(z^3) = \frac{1}{m\sqrt{m}}E(x-m)^3 = \frac{1}{\sqrt{m}}$$

$$\mu_4 \text{ for } z = E(z^4) = \frac{1}{m^2}E(x-m)^4 = \frac{1}{m}+3$$

As $m \to \infty$, $z \to a$ normal variate.

$$\therefore \quad \mu_2 \text{ for normal variate} = \underset{m\to\infty}{\text{Lt}}\ \mu_2 \text{ for } z = 1$$

$$\therefore \quad \mu_3 \text{ for normal variate} = \underset{m\to\infty}{\text{Lt}}\ \mu_3 \text{ for } z = \underset{m\to\infty}{\text{Lt}}\ \frac{1}{\sqrt{m}} = 0$$

$$\therefore \quad \mu_4 \text{ for normal variate} = \underset{m\to\infty}{\text{Lt}}\left(3+\frac{1}{m}\right) = 3.$$

Ex. 10-60. *For a certain normal distribution the first moment about 10 is 40 and that the 4th moment about 50 is 48, what is the A.M. and s.d. of the dist.?*

Sol. Let m and σ be A.M. and *s.d.*

Now $$\mu_1'(10) = 40$$

$$\therefore \quad E(x-10) = 40$$

or $$E(x) = 50$$

$$\therefore \quad m = 50$$

Also $$\mu_4 = 48$$

$$\therefore \quad 3\sigma^4 = 48$$

$$\therefore \quad \sigma = 2.$$

Ex. 10-61. *If X is a normal variate with mean 30 and s.d. 5. Find the probabilities that*

(*i*) $26 \le X \le 40$, (*ii*) $|X-30| > 5$.

Sol. (*i*) $P\{26 \le X \le 40\} = P\{26 \le X \le 30\} + P\{30 \le X \le 40\}$

Put $$Z = \frac{X-30}{5}$$

(*ii*)

Ex. 10-62. *For a normal dist*

variate such that the probability

Sol. Let X be normal variate.

Then dist. of X is

$P(2 < X <$

Put $\frac{X-}{3}$

$\therefore$ 0·41

$\therefore$ $x-$

$\therefore$

Ex. 10-63. *Prove that, for the and the s.d are approximately in*

Sol. Let Q_1 and Q_3 be the

$$= P\{-0\cdot 8 \le Z \le 0\} + P\{0 \le Z \le 2\}$$

$$= P\{0 \le Z \le 0\cdot 8\} + P\{0 \le Z \le 2\}$$

$$= 0\cdot 2881 + 0\cdot 4772 = 0\cdot 7653.$$

(using normal tables)

(*ii*) $$P\{|X-30| > 5\} = 1 - P\{|X-30| \le 5\}$$

$$= 1 - P\{25 \le x \le 35\}$$

$$= 1 - 2P\{30 \le X \le 35\}$$

$$= 1 - 2P\{0 \le Z \le 1\}$$

$$= 1 - 2(0\cdot 3413)$$

$$= 0\cdot 3174.$$

Ex. 10-62. *For a normal distribution with mean 2 and variance 9, find the value x of the variate such that the probability of the variate lying in the interval* $(2, x)$ *is* $0\cdot 4115$.

Sol. Let X be normal variate.

Then dist. of X is

$$dP = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(X-2)}{3}\right)^2} dx$$

$$P(2 < X < x) = \frac{1}{3\sqrt{2\pi}} \int_2^x e^{-\frac{1}{2}\left(\frac{X-2}{3}\right)^2} dX$$

Put $$\frac{X-2}{3} = z$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{\frac{x-2}{3}} e^{-\frac{1}{2}z^2} dz$$

$$\therefore \quad 0\cdot 4115 = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{x-2}{3}} e^{-\frac{1}{2}z^2} dz$$

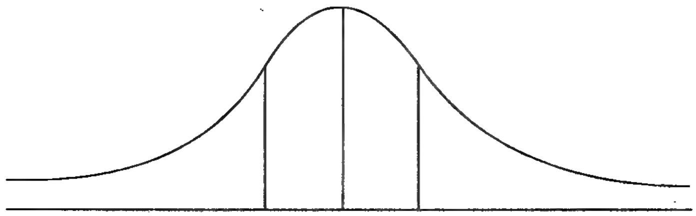
$$\therefore \quad \frac{x-2}{3} = 1\cdot 35$$

$$\therefore \quad x = 2 + 4\cdot 05 = 6\cdot 05.$$

Ex. 10-63. *Prove that, for the normal distribution the quartile deviation, mean deviation and the s.d are approximately in the ratio 10:12:15.*

Sol. Let Q_1 and Q_3 be the Quartiles

Then $$P\{x \le Q_1\} = 0\cdot 25$$



$$\therefore \quad P\{Q_1 < x < m\} = 0\cdot 5 - 0\cdot 25$$
$$= 0\cdot 25$$

$$\therefore \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{Q_1}^{m} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dz = 0\cdot 25$$

Put $$\frac{m-x}{\sigma} = y$$

$$\therefore \quad \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{m-Q_1}{\sigma}} e^{-\frac{1}{2}y^2} dy = 0\cdot 25$$

$$\therefore \quad \frac{m-Q_1}{\sigma} = 0\cdot 6744 \qquad \ldots(1)$$

Also $$P(x \ge Q_3\} = 0\cdot 25$$

$$\therefore \quad P\{m \le x < Q_3\} = 0\cdot 25$$

$$\therefore \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{Q_3} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = 0\cdot 25$$

or $$\frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{Q_3-m}{\sigma}} e^{-\frac{1}{2}y^2} dy = 0\cdot 25$$

$$\therefore \quad \frac{Q_3-m}{\sigma} = 0\cdot 6744 \qquad \ldots(2)$$

From (1) and (2)

$$\frac{Q_3-Q_1}{2} = 0\cdot 6744\,\sigma \simeq \frac{2}{3}\sigma$$

$$\therefore \text{ Quartile Deviation } = \frac{2}{3}\sigma$$

Also Mean deviation

$\therefore$ Q.D.

Ex. 10-64. *If two normal i of universe A is k times that of is $\frac{1}{k}$ times that of universe B.*

Sol. Let N be the total fre

Then

Let m_1 and m_2 be the A. I The frequency functions c

and

Evidently $F_A(x)$ is max f

$[F_A($

Similarly $[F_B($

$\therefore$ $\frac{[F_A(x}{[F_B(x}$

$\therefore$ $[F_A($

Ex. 10-65. *Assume the me 10·8 (in)^2. How many soldier tall ? (Given that the area und 0·1368 and between x = 0 and*

Sol. Let x inches be the h

Then x is a normal variate
$\therefore$ Dist. of x is

Also Mean deviation $= \frac{4}{5}\sigma$

$\therefore$ Q.D. : M.D. : S.D. : : 10 : 12 : 15.

Ex. 10-64. *If two normal universes A and B have the same total frequency but the s.d. of universe A is k times that of the universe B, show that maximum frequency of universe A is $\frac{1}{k}$ times that of universe B.*

Sol. Let N be the total frequency and σ_1, σ_2 be the s.d. of A and B.

Then $\sigma_1 = k\sigma_2$

Let m_1 and m_2 be the A. Ms. of A and B.

The frequency functions of A and B are

$$F_A(x) = \frac{N}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m_1}{\sigma_1}\right)^2}$$

and
$$F_B(x) = \frac{N}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m_2}{\sigma_2}\right)^2}$$

Evidently $F_A(x)$ is max for $x = m_1$.

$$[F_A(x)]_{\max} = \frac{N}{\sigma_1\sqrt{2\pi}}$$

Similarly
$$[F_B(x)]_{\max} = \frac{N}{\sigma_2\sqrt{2\pi}}$$

$$\therefore \quad \frac{[F_A(x)]_{\max}}{[F_B(x)]_{\max}} = \frac{\sigma_2}{\sigma_1} = \frac{1}{k}$$

$$\therefore \quad [F_A(x)]_{\max} = \frac{1}{k}[F_B(x)]_{\max}.$$

Ex. 10-65. *Assume the mean heights of soldiers to be 68·22 inches with a variance of 10·8 (in)^2. How many soldiers in a regiment of 1000 would you expect to be over 6 feet tall ? (Given that the area under the standard normal curve between x = 0 and x = 0·35 is 0·1368 and between x = 0 and x = 1·15 is 0·3746).*

Sol. Let x inches be the height.

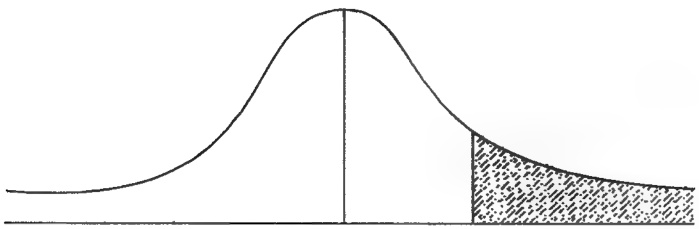
Then x is a normal variate with mean 68·22 inches and variance 10·8 (in)^2.

$\therefore$ Dist. of x is

$$dP = \frac{1}{\sqrt{10\cdot8}\,\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-68\cdot22}{\sqrt{10\cdot8}}\right)^2} dx$$

$$\therefore \qquad P\{x > 72\} = 0\cdot5 - \int_{x=68\cdot22}^{72} \frac{1}{\sqrt{10\cdot8}\,\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-68\cdot22}{\sqrt{10\cdot8}}\right)^2} dx$$

Put $$z = \frac{x - 68\cdot22}{\sqrt{10\cdot8}}$$



$$\therefore \qquad P\{x > 72\} = 0\cdot5 - \frac{1}{\sqrt{2\pi}} \int_0^{1\cdot15} e^{-\frac{1}{2}z^2} dz$$

$$= 0\cdot5 - (0\cdot3746) \qquad \text{(given)}$$

$$= 0\cdot1254.$$

$\therefore$ In a regiment of 1000, the number of soldiers taller than 6 feet.

$$= 1000 \times 0\cdot1254 = 125\cdot4$$

$$\approx 125.$$

Ex. 10-66. *If* $\log_{10} x$ *is normally distributed with mean 4 and variance 4, find the probability of* $1\cdot202 < x < 83180000$.

(Given $\log_{10}\ 1202 = 3\cdot08$, $\log_{10}\ 8318 = 3\cdot92$).

(*b*) $\log_{10}\ x$ *is normally distributed with mean 7 and variance 3.* $\log_{10} y$ *is normally distributed with mean 3 and unit variance. If the distribution of x and y are independent, find the prob. of*

$$1\cdot202 < \frac{x}{y} < 83180000.$$

Sol. (*a*) Let $y = \log_{10} x$.

Dist. of y is

$$dP = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left\{\frac{y-4}{2}\right\}^2} dy$$

Now $P\{1\cdot202 < x < 83180000\} = P\{0\cdot08 < y < 7\cdot92\}$

$$= \frac{1}{2\sqrt{2\pi}} \int_{0\cdot08}^{7\cdot22} e^{-\frac{1}{2}\left\{\frac{y-4}{2}\right\}^2} dy$$

Put

$P\{1\cdot202 < x < 83$

(*b*) Let $z_1 = \log_{10}\ x$ a

Then $z = z_1 - z_2$ is als

$\therefore$ Dist. of z is

Now

Ex. 10-67. *If the skulls is under 75, between 75 an normal) the mean and s.d. given that if*

then $f(0\cdot20) = 0\cdot08$ *a*

Sol. Let m and σ be th

Then dist. of x is

$$\frac{}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-68\cdot 22}{\sqrt{10\cdot 8}}\right)^2} dx$$

dz

(given)

than 6 feet.

5·4

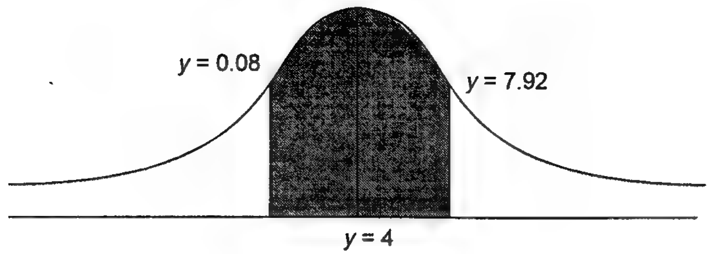
ean 4 and variance 4, find the

variance 3. $\log_{10} y$ *is normally*

on of x and y are independent,

dy

Put $\qquad \dfrac{y-4}{2} = z$



$$P\{1\cdot 202 < x < 83180000\} = \frac{1}{2\sqrt{2\pi}} \int_{-1\cdot 96}^{1\cdot 96} e^{-\frac{1}{2}z^2} dz$$

$$= 2.\frac{1}{\sqrt{2\pi}} \int_{0}^{1\cdot 96} e^{-\frac{1}{2}z^2} dz = 2(0\cdot 4750)$$

(from normal tables)

$$= 0\cdot 95.$$

(*b*) Let $z_1 = \log_{10} x$ and $z_2 = \log_{10} y$.

Then $z = z_1 - z_2$ is also a normal variate with mean $7-3=4$ and variance $3+1=4$.

$\therefore$ Dist. of z is

$$dP = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-4}{2}\right)^2} dt$$

Now $\qquad P\left\{1\cdot 202 < \dfrac{x}{y} < 83180000\right\}$

$$= P\{0\cdot 08 < z < 7\cdot 92\}$$

$$= 0\cdot 95 \qquad \{\text{from } (a)\}.$$

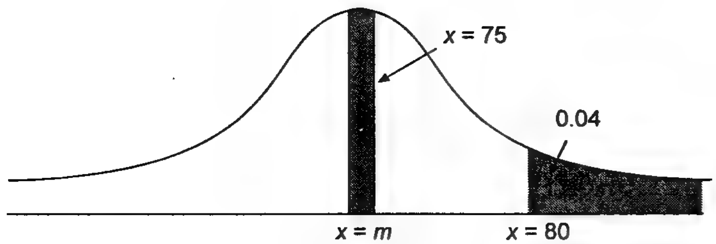
Ex. 10-67. *If the skulls are classified A, B and C according as the length breadth index is under 75, between 75 and 80 and over 80, find approximately (assuming that the dist. is normal) the mean and s.d. of a series in which A are 58% B are 38% and C are 4% being given that if*

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} dx$$

then $f(0\cdot 20) = 0\cdot 08$ *and* $f(1\cdot 75) = 0\cdot 46$.

Sol. Let m and σ be the mean and s.d. respectively and x be the length breadth index. Then dist. of x is

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx.$$



Now $P\{x < 75\} = 0 \cdot 58$ which is greater than 0·50 and hence the ordinate $x = 75$ is on the right of $x = m$.

From fig., $P\{m < x < 75\} = 0 \cdot 08$

$$\therefore \qquad \frac{1}{\sigma\sqrt{2\pi}} \int_m^{75} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = 0 \cdot 08$$

Put $$\frac{x-m}{\sigma} = z$$

$$\therefore \qquad \frac{1}{\sqrt{2\pi}} \int_0^{\frac{75-m}{\sigma}} e^{-\frac{1}{2}z^2} dz = 0 \cdot 08$$

$\therefore$ From given.

$$\frac{75-m}{\sigma} = 0 \cdot 20 \qquad \ldots(1)$$

Again $\qquad P(x > 80) = 0 \cdot 04$

$\therefore \qquad P\{m < x < 80\} = 0 \cdot 46$

$$\therefore \qquad \frac{1}{\sigma\sqrt{2\pi}} \int_m^{80} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = 0 \cdot 46$$

or $$\frac{1}{\sqrt{2\pi}} \int_m^{\frac{80-m}{\sigma}} e^{-\frac{1}{2}z^2} dz = 0 \cdot 46$$

$$\therefore \qquad \frac{80-m}{\sigma} = 1 \cdot 75 \qquad \text{(from given)} \qquad \ldots(2)$$

From (1) and (2)

$$m = 74 \cdot 4 \qquad \text{(approx.)}$$
$$\sigma = 3 \cdot 2 \qquad \text{(approx.)}$$

Ex. 10-68. *One thousand candidates in an examination were grouped into three classes I, II, III in descending order of merit. The numbers in the first two classes were 50 and 350 respectively. The highest and lowest marks in class II were 60 and 50 respectively. Assuming the distribution to be normal, prove that the average mark is 48·2 approximately and standard deviation 7·1 approximately. Given that :*

$\frac{x}{\sigma}$

0·2

0·3

0·4

where the area A is meas

Sol. Number of cand

$= 1000 -$

$\therefore$

Also

$\therefore \quad P\{n$

Put

$\therefore \quad P\{0 < z$

From given data

Value of A for $\frac{x}{\sigma} = ($

Value of A for $\frac{x}{\sigma} = ($

$\therefore$ Increment in A fo

$\therefore$ Increment in $\frac{x}{\sigma}$ 1

$\therefore$ Value of $\frac{x}{\sigma}$ (for

$\therefore$

Also

$P\{i$

$\therefore \quad P\{0 < z$

ce the ordinate $x = 75$ is on

...(1)

given) ...(2)

rox.)
rox.)

re grouped into three classes
wo classes were 50 and 350
id 50 respectively. Assuming
approximately and standard

$\frac{x}{\sigma}$	A	$\frac{x}{\sigma}$	A
0·2	0·079	1·5	0·433
0·3	0·118	1·6	0·445
0·4	0·155	1·7	0·455

where the area A is measured from the mean zero to any ordinate x.

Sol. Number of candidates getting III class.

$$= 1000 - (350 + 50) = 600$$

$$\therefore \quad P\{x < 50\} = 0 \cdot 6$$

Also $$P\{x < m\} = 0 \cdot 5$$

$$\therefore \quad P\{m < x < 50\} = 0 \cdot 1$$

Put $$z = \frac{x - m}{\sigma}$$

$$\therefore \quad P\left\{0 < z < \frac{50 - m}{\sigma}\right\} = 0 \cdot 1$$

From given data

Value of A for $\frac{x}{\sigma} = 0 \cdot 2$ is $0 \cdot 079$

Value of A for $\frac{x}{\sigma} = 0 \cdot 3$ is $0 \cdot 118$

$\therefore$ Increment in A for increment $0 \cdot 1$ in $\frac{x}{\sigma}$

$$= 0 \cdot 039.$$

$\therefore$ Increment in $\frac{x}{\sigma}$ for increment $0 \cdot 021$ in A

$$= \frac{0 \cdot 1}{0 \cdot 039} 0 \cdot 021 = 0 \cdot 054$$

$\therefore$ Value of $\frac{x}{\sigma}$ (for $A = 0 \cdot 1$) $= 0 \cdot 2 + 0 \cdot 054$

$$= 0 \cdot 254$$

$$\therefore \quad \frac{50 - m}{\sigma} = 0 \cdot 254 \qquad \ldots(1)$$

Also $$P(x > 60\} = 0 \cdot 05$$

$$P\{m < x < 60\} = 0 \cdot 5 - 0 \cdot 05 = 0 \cdot 45$$

$$\therefore \quad P\left\{0 < z < \frac{60 - m}{\sigma}\right\} = 0 \cdot 45$$

As above, from given data

$$\frac{60-m}{\sigma} = 1{\cdot}65 \qquad \ldots(2)$$

From (1) and (2)

$$\sigma = 7{\cdot}1 \quad \text{(approx.)}$$

and

$$m = 48{\cdot}2 \text{ (approx.)}$$

Ex. 10-69. *In a normal dist. 31% of the items are under 45 and 8% are over 64. Find the mean and s.d. of the distribution.*

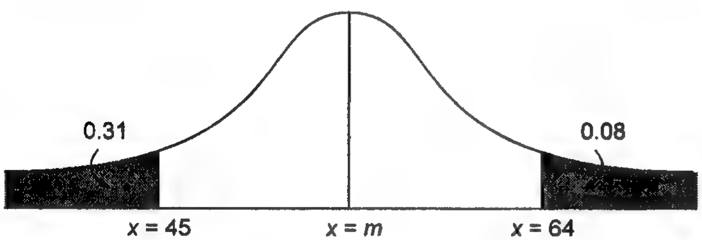
Sol. Let m be the mean and σ the s.d. Then

$$P\{x < 45\} = 0{\cdot}31$$

$$\therefore \quad P\{45 < x < m\} = 0{\cdot}19$$

or

$$P\left\{\frac{45-m}{\sigma} < z < 0\right\} = 0{\cdot}19$$



$$\therefore \quad P\left\{0 < z < \frac{m-45}{\sigma}\right\} = 0{\cdot}19$$

$$\therefore \quad \frac{m-45}{\sigma} = 0{\cdot}496 \qquad \ldots(1)$$

Similarly

$$P\left\{0 < z < \frac{64-m}{\sigma}\right\} = 0{\cdot}42$$

$$\therefore \quad \frac{64-m}{\sigma} = 1{\cdot}405 \qquad \ldots(2)$$

From (1) and (2)

$$m = 10 \text{ (approx.)}$$
$$\sigma = 50 \text{ (approx.)}$$

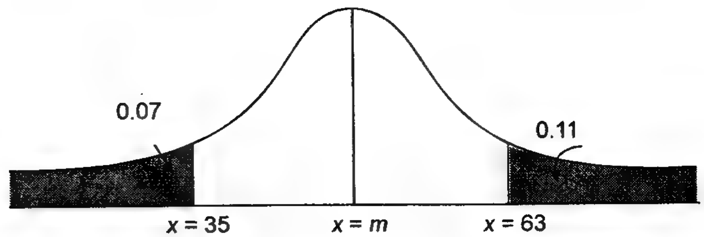
Ex. 10-70. *In a distribution exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and s.d. of the dist. ?*

Sol.

$$P\{x < 35\} = 0{\cdot}07$$

$$\therefore \quad P\{35 < x < m\} = 0{\cdot}43$$

$$\therefore \quad P\left\{0 < z < \frac{m-35}{\sigma}\right\} = 0{\cdot}43$$



$$\therefore \qquad \frac{m-35}{\sigma} = 1{\cdot}476 \qquad \ldots(1)$$

$$P\{x < 63\} = 0{\cdot}89$$

$$\therefore \qquad P\{m < x < 63\} = 0{\cdot}39$$

$$\therefore \qquad P\left\{0 < z < \frac{63-m}{\sigma}\right\} = 0{\cdot}39$$

$$\therefore \qquad \frac{63-m}{\sigma} = 1{\cdot}226. \qquad \ldots(2)$$

From (1) and (2)

$$\sigma = 10{\cdot}36 \text{ (approx.)}$$
$$m = 50{\cdot}29 \text{ (approx.)}$$

Ex. 10-71. *Five thousand candidates appeared in a certain examination paper carrying a maximum of 100 marks. It was found that the marks were normally distributed with mean 39·5 and s.d. 12·5. Determine approximately the number of students who secured a first class for which a minimum of 60 marks is necessary you may use the table given below :*

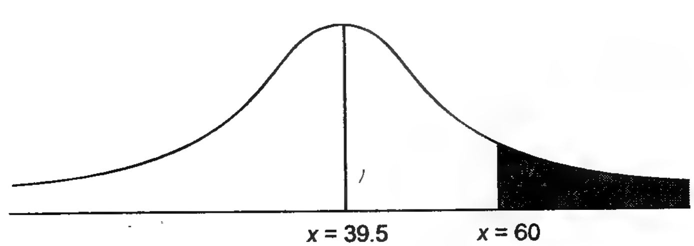
The proportion A of the whole area of the normal curve lying to the left of the ordinate at the deviation $\frac{x}{\sigma}$ *is*

$\frac{x}{\sigma}$	1·5	1·6	1·7	1·8
A :	0·93319	0·94520	0·95543	0·95407

Sol. $P\{39{\cdot}5 < x < 60\}$

$$= P\{0 < z < 1{\cdot}64\} = 0{\cdot}94929 - 0{\cdot}5$$
$$= 0{\cdot}44929$$

$$\therefore \qquad P\{x > 60\} = 0{\cdot}5 - 0{\cdot}44929$$
$$= 0{\cdot}05071$$



$\therefore$ **No. of students getting first class = 5000 × 0·05071 = 253.**

Ex. 10-72. *A minimum height is to be prescribed for eligibility to government services such that 60% of the youngmen will have a fair chance of coming upto that standard. The heights of youngmen are normally distributed with mean 60·6″ and s.d. 2·55″. Determine the minimum specification.*

$$\left\{\textit{From table if } f(t)=\frac{1}{\sqrt{2\pi}}\int^{\infty}\exp\left(-\frac{t^2}{2}\right)dt, \textit{ then } f(-0\cdot2533)=0\cdot6\right\}$$

Sol. Let h be the minimum height prescribed.

Then $$\frac{1}{\sigma\sqrt{2\pi}}\int_h^{\infty} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}dx = 0\cdot6$$

Put $$t = \frac{x-m}{\sigma}$$

$\therefore$ $$\frac{1}{\sqrt{2\pi}}\int_{\frac{h-m}{\sigma}}^{\infty} e^{-\frac{1}{2}t^2}dt = 0\cdot6$$

$\therefore$ $$\frac{h-m}{\sigma} = -0\cdot2533$$

Here $m = 60\cdot6,\qquad \sigma = 2\cdot55$

$\therefore$ $h = 59\cdot95 \simeq 60.$

Ex. 10-73. *The local authorities in a certain city installed 2,000 electric lamps in streets. If the lamps have an average life of 1,000 burning hours with a s.d., of 200 hours.*

(*a*) *What number of lamps might be expected to fail in first 700 burning hours ?*

(*b*) *After what period of burning hours would you expect that 10% of the lamps would have failed ?*

Assume that lives of the lamps are normally distributed.

Given that if $$F(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}t^2}dt.$$

Then $F(1\cdot50) = 0\cdot933$

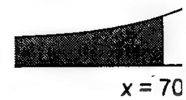
and $F(1\cdot28) = 0\cdot900.$

Sol. Let x hours be the life of a lamp.

(*a*) Since normal curve is symmetrical about $x = 1000$

$$\begin{aligned} P\{x<700\} &= P\{x>1300\} \\ &= 1-P\{x<1300\} \\ &= 1-\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{1300} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}dx. \end{aligned}$$

Here



Put $z =$

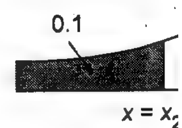
$\therefore$ $P\{x$

$\therefore$ Number of lamps exp

(*b*) Let $x = x_1$ be s.t.

$P\{$

$\therefore$ $P\{$



or

or

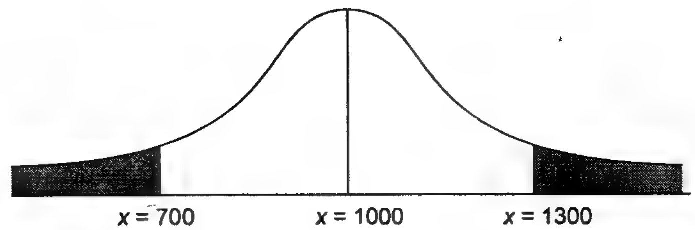
$\therefore$

$\therefore$

Let

$P\{$

Here $m = 1000, \sigma = 200$



Put $$z = \frac{x-m}{\sigma} = \frac{x-1000}{200}$$

$\therefore$ $$P\{x < 700\} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1\cdot 5} e^{-\frac{1}{2}z^2}\, dz$$

$$= 1 - 0\cdot 933 = 0\cdot 067$$

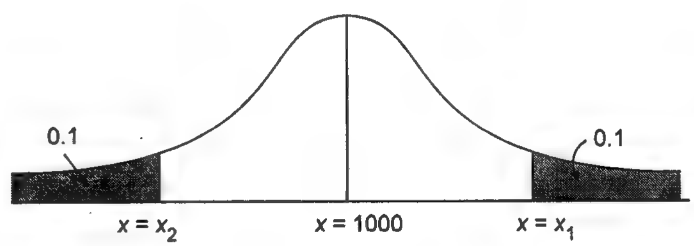
$\therefore$ Number of lamps expected to fail in first 700 hours of burning

$$= 2000 \times 0\cdot 067 = 134$$

(*b*) Let $x = x_1$ be s.t.

$$P\{x > x_1\} = 0\cdot 1$$

$\therefore$ $$P\{x < x_1\} = 0\cdot 9$$



or $$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = 0\cdot 9$$

or $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x_1-m}{\sigma}} e^{-\frac{1}{2}z^2}\, dz = 0\cdot 9$$

$\therefore$ $$\frac{x_1 - m}{\sigma} = 1\cdot 28$$

$\therefore$ $$x_1 = 1000 + 200\,(1\cdot 28) = 1256$$

Let $x = x_2$ be s.t.

$$P\{x < x_2\} = 0\cdot 1$$

Then by symmetry of normal curve about $x = 1000$,

$$x_2 = 1000 - 256 = 744$$

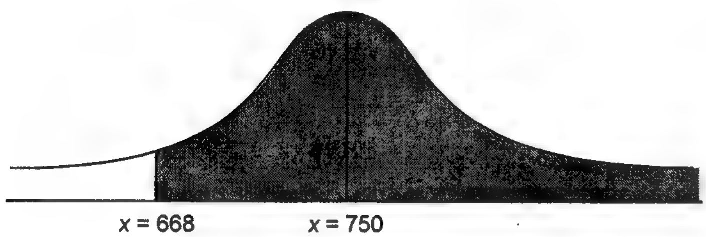
$\therefore$ After 744 hours of burning, 10% lamps are expected to fail.

Ex. 10-74. *The incomes of a group of 10,000 persons were found to be normally distributed with mean = Rs. 750 p.m. and s.d. = Rs. 50. Show that of this group about 95% had income exceeding Rs. 668 and only 5% had income exceeding Rs. 832. What was the lowest income among the richest 100 ?*

Sol. Let x be the income variate

Here $\quad m = 750, \quad \sigma = 50$

(*i*)
$$\begin{aligned} P\{x > 668\} &= 0\cdot5 + P(668 < x < 750\} \\ &= 0\cdot5 + P\{-1\cdot64 < z < 0\} \\ &= 0\cdot5 + P\{0 < z < 1\cdot64\} \\ &= 0\cdot5 + 0\cdot4495 \\ &= 0\cdot9495 \end{aligned}$$



$\therefore$ Percentage of persons having income exceeding Rs. 668

$$= 94\cdot95 \simeq 95\%$$

(*ii*)
$$\begin{aligned} P\{x > 832\} &= 0\cdot5 - P(750 < x < 832\} \\ &= 0\cdot5 - P\{0 < z < 1\cdot64\} \\ &= 0\cdot5 - 0\cdot4495 = 0\cdot0505 \end{aligned}$$

$\therefore$ Percentage of persons having income exceeding Rs. 832

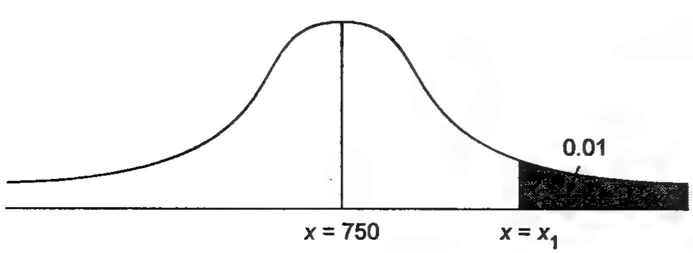
$$= 5\%$$

(*iii*) Let $x = x_1$ be s.t.

$$P\{x > x_1\} = 0\cdot01$$

Then $x = x_1$ is the lowest income among the richest 100.

$$\therefore \quad P\{750 < x < x_1) = 0\cdot49$$



or
$$P\left\{0 < z < \frac{x_1 - 750}{50}\right\} = 0\cdot49.$$

x_1

$\therefore$

$\therefore$

Ex. 10-75. *In a certain*
obtained were 50% and the s
than 60 marks, supposing the
normal curve from x = 0 to x

Sol. Let x denote the ma

Then x is $N(50, 5)$

Now $\quad P($

Put

$\therefore \quad P($

$\therefore$ Expected number of

Ex. 10-76. *If x and y are*
16 respectively, determine λ

$P(2x +$

Sol. Let $u = 2x + y, v =$

Then $\quad u$ is a N

mean =

and variance $= 4\sigma_x^2 + \sigma_y^2$

$= 4.9 +$

and v is a N.V. with mean

$\bar{v} = 4\bar{x}$

and variance $= 16\sigma_x^2 + 9\sigma_y$

$= 16.9$

Now $\quad P(2x +$

Put

$$\therefore \qquad \frac{x_1 - 750}{50} = 2{\cdot}3267$$

$$\therefore \qquad x_1 = 866{\cdot}34.$$

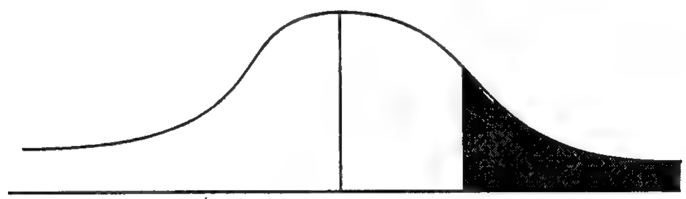
Ex. 10-75. *In a certain examination 2000 students appeared. The average marks obtained were 50% and the s.d. was 5%. How many students do you expect to obtain more than 60 marks, supposing the marks to be distributed normally? (Area under the standard normal curve from x = 0 to x = 2, is 0·4772).*

Sol. Let x denote the marks obtained.

Then x is $N(50, 5)$

Now $\qquad P(x > 60) = 0{\cdot}5 - P\{50 < x < 60\}$

Put $\qquad z = \dfrac{x - 50}{5}$



$$\therefore \qquad P(x > 60) = 0{\cdot}5 - P(0 < z < 2)$$
$$= 0{\cdot}5 - 0{\cdot}4772 = 0{\cdot}0228$$

$\therefore$ Expected number of students getting more than 60% marks

$$= 2000\,(0{\cdot}0228)$$
$$= 45{\cdot}6 \simeq 46.$$

Ex. 10-76. *If x and y are independent normal variates with means 6, 7 and variances 9, 16 respectively, determine λ such that*

$$P(2x + y \le \lambda) = P(4x - 3y \ge 4\lambda)$$

Sol. Let $u = 2x + y, v = 4x - 3y$

Then $\qquad u$ is a *N.V.* with

$\qquad$ mean $= 2\bar{x} + \bar{y} = 2.6 + 7 = 19$

and variance $= 4\sigma_x^{\,2} + \sigma_y^{\,2}$

$\qquad = 4.9 + 16 = 52$

and v is a N.V. with mean

$\qquad \bar{v} = 4\bar{x} - 3\bar{y} = 4.6 - 3.7 = 3$

and variance $= 16\sigma_x^{\,2} + 9\sigma_y^{\,2}$

$\qquad = 16.9 + 9.16 = 288.$

Now $\qquad P(2x + y \le \lambda) = P(u \le \lambda)$

Put $\qquad z_1 = \dfrac{u - 19}{\sqrt{52}}$

Then z_1 is $N(0, 1)$ and

$$\Rightarrow \qquad u = 19 + \sqrt{52}\, z_1$$

$$\therefore \qquad P(2x + y \le \lambda) = P\{19 + \sqrt{52}\, z_1 \le \lambda\}$$

$$= P\left\{z_1 \le \frac{\lambda - 19}{\sqrt{52}}\right\} \qquad \ldots(1)$$

Similarly $\quad P(4x - 3y \ge 4\lambda) = P(v \ge 4\lambda)$

Put $$z_2 = \frac{v - 3}{\sqrt{288}}$$

Then z_2 is $N(0, 1)$ and

$$v = 3 + \sqrt{288}\, z_2$$

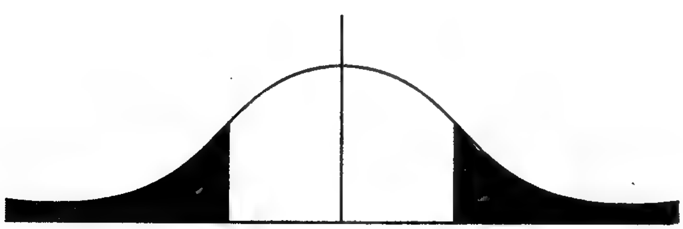
$$\therefore \qquad P(4x - 3y \ge 4\lambda) = P\left\{z_2 \ge \frac{4\lambda - 3}{\sqrt{288}}\right\} \qquad \ldots(2)$$

$$\therefore \text{ Given Eq.} \Rightarrow \quad P\left\{z_1 \le \frac{\lambda - 19}{\sqrt{52}}\right\} = P\left\{z_2 \ge \frac{4\lambda - 3}{\sqrt{288}}\right\}$$

Since z_1, z_2 both are $N(0, 1)$, we can replace both of them by z (which is also $N(0, 1)$.

$$\therefore \qquad P\left\{z \le \frac{\lambda - 19}{\sqrt{52}}\right\} = P\left\{z \ge \frac{4\lambda - 3}{\sqrt{288}}\right\}$$

$\Rightarrow \dfrac{\lambda - 19}{\sqrt{52}}$ and $\dfrac{4\lambda - 3}{\sqrt{288}}$ must be on the opposite sides of $z = 0$ (which is mean of z) and are equidistant from it.



$$\therefore \qquad \frac{\lambda - 19}{\sqrt{52}} = -\frac{4\lambda - 3}{\sqrt{288}}$$

$$\Rightarrow \qquad \lambda = \frac{114\sqrt{2} + 3\sqrt{13}}{6\sqrt{2} + 4\sqrt{13}} = 7{\cdot}51.$$

Ex. 10-77. *If x is a N* (2, 3

Sol.

Put

$\therefore$ $P\Big(y$

Ex. 10-78. *If* $N(r) = P(x$

Sol. Since normal curve is

$P(0 \le x$

$x = -r$

Now

Ex. 10-77. *If x is a N (2, 3), find* $P\left[y \geq \frac{3}{2}\right]$ *where* $y = x - 1$.

Sol. Now $P\left(y \geq \frac{3}{2}\right)$

$$= P\left(x - 1 \geq \frac{3}{2}\right)$$

$$= P(x \geq 2 \cdot 5)$$

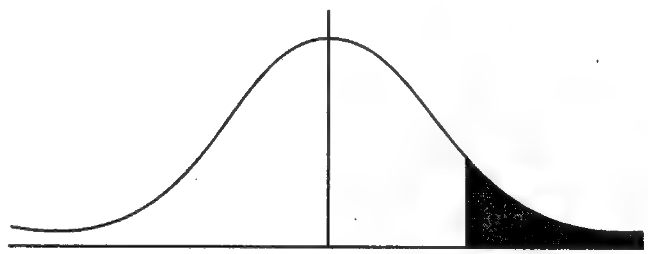
$$= 0 \cdot 5 - P(2 \leq x \leq 2 \cdot 5)$$

Put $z = \dfrac{x-2}{3}$

$\therefore$ $P\left(y \geq \frac{3}{2}\right) = 0 \cdot 5 - P(0 \leq z \leq 0 \cdot 17)$

$$= 0 \cdot 5 - \cdot 0675$$

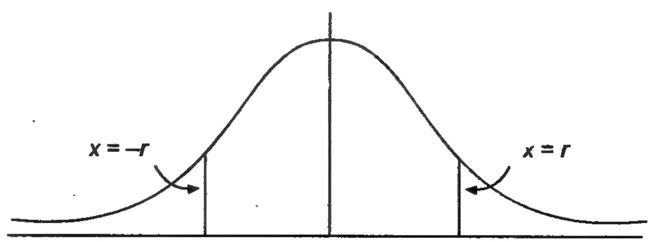
$$= 0 \cdot 4325$$



Ex. 10-78. *If* $N(r) = P(x \leq r)$, *where x is a N (0, 1), show that* $N(-r) = 1 - N(r)$.

Sol. Since normal curve is symmetrical about $x = 0$,

$$P(0 \leq x \leq r) = P(-r \leq x \leq 0) \qquad \ldots(1)$$



Now $N(-r) = P(x \leq -r)$

$$= 0 \cdot 5 - P(-r \leq x \leq 0)$$

$$= 0 \cdot 5 - P(0 \leq x \leq r)$$

$$= 1 - \{0 \cdot 5 + P(0 \leq x \leq r)$$

$$= 1 - P(x \le r)$$
$$= 1 - N(r).$$

Ex. 10-79. *Let x be normally distributed with mean* $\mu(>0)$ *and s.d.* σ. *Suppose* σ^2 *is some function of* μ *say* $\sigma^2 = h(\mu)$. *Choose* $h(.)$ *so that* $P(x \le 0)$ *does not depend on* μ.

Sol. Take $$z = \frac{x-\mu}{\sigma}$$

$$P(x \le 0) = P\left(z \le -\frac{\mu}{\sigma}\right)$$

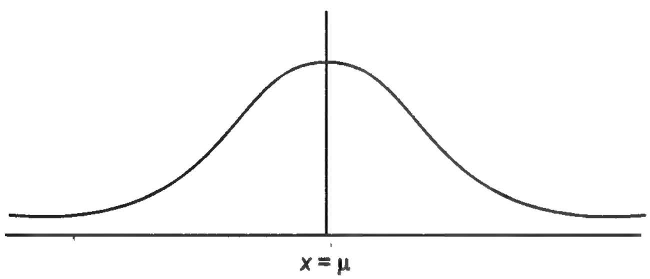
for this to be independent of μ, take

$$\sigma = \mu.$$

Then $$P(x \le 0) = P(z \le -1) = \Phi(-1).$$

Ex. 10-80. *If x is normally distributed with mean* $\mu > 0$ *and variance* $\sigma^2 = \mu^2$, *evaluate* $P(x < -\mu / x < \mu)$.

Sol. $P\{(x < -\mu) \cap (x < \mu)\}$



$$= P(x < \mu)\, P\{x < -\mu / x < \mu\}$$

$$P(x < -\mu) = 0 \cdot 5 P(x < -\mu / x < \mu)$$

$$\{\because (x < -\mu) \cap (x < \mu) = (x < -\mu)\}$$

$\therefore$ $$P(x < -\mu / x < \mu) = 2\, P(x < -\mu)$$

Put $$Z = \frac{x-\mu}{\sigma} = \frac{x-\mu}{\mu}$$

$$= 2P(z < -2)$$
$$= 2P(z > 2) = 2\{0 \cdot 5 - P(0 \le z \le 2)\}$$
$$= 2\{0 \cdot 5 - 0 \cdot 4772\}$$
$$= 0 \cdot 0456$$

1. If x is normally distributed w
2. Show that the s.d. of a norm deviation about the mean.
3. If x is normally distributed w of the standard normal cumu
4. x is a normal variate with mea
   (*i*) Values of the probability $x = -\infty, 46 \cdot 6275, 50, 53 \cdot 37$
   (*ii*) Probabilities over the in $(-\infty, 46 \cdot 6275), (46 \cdot 6275, 5$
   Comment on the various res
   (Given that $\frac{1}{\sqrt{2\pi}} = 0 \cdot 3989$
5. x is a normal variate with pı $f(x) = 0 \cdot 3989 \exp\{-0 \cdot 00$
   Express $f(x)$ in the standar of x.
6. x is a normal variate with pı $f(x) = 0 \cdot 7978 \exp.\{-2x^2 +$
   Express $f(x)$ in standard f the distribution of x.
7. Assuming a Normal distribu
   (*i*) the number of observa
   (*ii*) the value of the variat
   [**Hint :** for (*ii*) if r be the va

P

$\Rightarrow$ P

Put z

$\Rightarrow$ P

8

$\Rightarrow$

$\Rightarrow$ r

EXERCISES

1. If x is normally distributed with mean 2 and variance 2, find $P\{|x-1| \le 2\}$.

   [Ans. $0{\cdot}7624$]

2. Show that the s.d. of a normal distribution is approximately 20% more than the mean deviation about the mean.

3. If x is normally distributed with mean 2 and variance 2, express $P\{|x-1| \le 2\}$ in terms of the standard normal cumulative distribution function.

4. x is a normal variate with mean 50 and variance 25. Set out in tabular form the following :

   (*i*) Values of the probability density function for

   $x = -\infty, 46\cdot6275, 50, 53\cdot3725$ and $+\infty$.

   (*ii*) Probabilities over the intervals

   $(-\infty, 46\cdot6275), (46\cdot6275, 50), (50, 53\cdot3725)$ and $(53\cdot3725, +\infty)$.

   Comment on the various results that you obtain.

   $$\left(\text{Given that } \frac{1}{\sqrt{2\pi}} = 0\cdot3989, \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(0\cdot6745)^2} = 0\cdot31854\right)$$

5. x is a normal variate with probability density function

   $f(x) = 0\cdot3989 \exp\{-0\cdot005x^2 + 0\cdot5x - 12\cdot5\}$

   Express $f(x)$ in the standard form and hence or otherwise find the mean and variance of x.

6. x is a normal variate with probability density function

   $f(x) = 0\cdot7978 \exp.\{-2x^2 + 4x - 2\}$

   Express $f(x)$ in standard form and hence or otherwise find the mean and variance of the distribution of x.

7. Assuming a Normal distribution with $N = 1000; \mu = 80; \sigma = 15,$ find

   (*i*) the number of observations expected to lie between 65 and 110.

   (*ii*) the value of the variate beyond which 100% of the items lie.

   [**Hint :** for (*ii*) if r be the value, then

   $$P(x \ge r) = 1$$

   $$\Rightarrow \quad P(r \le x \le 80\} = 0\cdot5$$

   $$\text{Put} \quad z = \frac{80-x}{15}$$

   $$\Rightarrow \quad P\left\{0 \le z \le \frac{80-r}{15}\right\} = 0\cdot5$$

   $$\Rightarrow \quad \frac{80-r}{15} = 3\cdot92$$

   $$\Rightarrow \quad r = 21\cdot2$$

   [Ans. 818]

8. Express as an integral the probability that a normal variate with mean 5 and s.d. 2 would be observed between 2 and 3.

9. The following table gives frequencies of occurrence of a variable x between certain limits :

Variable x	Frequency
Less than 40	30
40 or more but less than 50	33
50 and more	37

The distribution is exactly normal. Find the distribution and also obtain the frequencies between $x = 50$ and $x = 60$.

[**Ans.** 11·68, 46·125, 25]

10. In a certain examination the percentages of passes and distinction were 45 and 10 respectively. Estimate the average marks obtained by the candidates, the minimum pass and distinction marks being 40 and 75 respectively. (Assume the distribution of marks to the normal).

[**Ans.** 36·1]

11. The marks obtained in a certain paper are found to be normally distributed. If 12·5% of the candidates obtain 60% or more marks, 39% obtain less than 30 marks, find the mean number of marks obtained by the candidates. Given

$\frac{x}{\sigma}$	0·27	0·28	0·29	1·14	1·15	1·6
A	0·6064	0·6102	0·6141	0·8727	0·8749	0·8770

[**Ans.** 36]

12. The height measurements of 600 adult males are arranged in ascending order and it is observed that the 180th and 450th measurements are 64·2" and 67·8" respectively. Assuming that the sample of heights is drawn from a normal population, estimate the mean and the s.d. of the population.

13. Steel rods are manufactured to be 3 inches in diameter but they are acceptable if they are inside the limits 2·99 inches and 3·01 inches. It is observed that 5% are rejected under size. Assuming that the diameters are normally distributed, find the s.d. of the distribution. Hence calculate what proportion of rejects would be if the permissible limits were widened to 2·985 inches and 3·015 inches.

[**Ans.** 1·36%]

14. In an examination it is laid down that a student passes if he secures 30% or more marks. He is placed in the first, second or third division according as he secures 60% or more marks, between 45% and 60% and marks between 30% and 45% respectively. He gets a distinction in case he secures 80% or more marks. It is noticed from the results that 10% of the students failed in the examination, whereas 5% of them obtained distinction. Calculate the percentage of students placed in the second division. (Assume marks to be distributed normally).

[**Ans.** 34%]

15. If x is a normal variate with mean 50 and s.d. 10, find $P(y \le 3137)$ where $y = x^2 + 1$.

[**Hint.** $P(y \le 3137) = P(x^2 + 1 \le 3137)$

$$= P(x^2 \le 3136)$$

$$= P(|x| \le 56)$$

Put $z = \frac{x-50}{10} \Rightarrow x = 50 +$

∴ $P(y \le$

$= P\{-$

$= P\{-$

$= P\{-$

$= 0\cdot$

$= 0\cdot$

10·4. Discrete Uniform Distrib

This distribution is of the f

where n is a positive integer. x i

Mean and Variance

μ

∴

M.G.F.

variate with mean 5 and s.d. 2

of a variable x between certain

ı and also obtain the frequencies

[**Ans.** 11·68, 46·125, 25]

nd distinction were 45 and 10
/ the candidates, the minimum
:ly. (Assume the distribution of

[**Ans.** 36·1]

normally distributed. If 12·5%
ain less than 30 marks, find the
ven

⁞	1·15	1·6
:7	0·8749	0·8770

[**Ans.** 36]

ged in ascending order and it is
64·2″ and 67·8″ respectively.
ormal population, estimate the

but they are acceptable if they
observed that 5% are rejected
distributed, find the s.d. of the
ts would be if the permissible

[**Ans.** 1·36%]

:s if he secures 30% or more
ı according as he secures 60%
en 30% and 45% respectively.
marks. It is noticed from the
, whereas 5% of them obtained
ı the second division. (Assume

[**Ans.** 34%]

$(y \le 3137)$ where $y = x^2 + 1$.

Put $z = \dfrac{x-50}{10} \Rightarrow x = 50 + 10z$

$$\begin{aligned}
\therefore \quad P(y \le 3137) &= P(-56 < x < 56) \\
&= P\{-56 < 50 + 10z < 56\} \\
&= P\{-10\cdot6 < z < 0\cdot6\} \\
&= P\{-10\cdot6 < z < 0\} + P\{0 < z < 0\cdot6) \\
&= 0\cdot5 + 0\cdot2258 \\
&= 0\cdot7258.
\end{aligned}$$

10·4. Discrete Uniform Distribution

This distribution is of the form

$$P(x) = \frac{1}{n+1},\ x = 0, 1, \ldots\ldots n$$

where n is a positive integer. x is called discrete **uniform random variable.**

Mean and Variance

$$\begin{aligned}
\bar{x} &= \frac{1}{n+1}\{0+1+\ldots\ldots+n\} = \frac{n(n+1)}{2(n+1)} \\
&= \frac{n}{2} \\
\mu_2'(0) &= \frac{1}{n+1}\{0^2+1^2+\ldots\ldots+n^2\} \\
&= \frac{n(n+1)(2n+1)}{6(n+1)} \\
&= \frac{n(2n+1)}{6} \\
\therefore \quad \mu_2 &= \mu_2'(0) - \bar{x}^2 \\
&= \frac{n(2n+1)}{6} - \frac{n^2}{4} \\
&= \frac{n(n+2)}{12}
\end{aligned}$$

M.G.F.

$$\begin{aligned}
M_0(t) &= E(e^{tx}) \\
&= \frac{1}{n+1}\sum_{x=0}^{n} e^{tx} \\
&= \frac{1}{n+1}\{1 + e^t + e^{2t} + \ldots + e^{nt}\}
\end{aligned}$$

$$= \frac{1}{n+1} \cdot \frac{1-e^{(n+1)t}}{1-e^t}.$$

10·5. Geometric Distribution

The prob. dist.

$P(x) = q^x p, \quad x = 0, 1, 2, \ldots\ldots, q = 1-p$

is called geometric distribution.

$$\bar{x} = E(x) = \sum_{x=0}^{\infty} x q^x p = p\{q + 2q^2 + 3q^3 + \ldots\}$$

$$= pq\{1 + 2q + 3q^2 + \ldots\ldots\} = pq(1-q)^{-2}$$

$$= \frac{q}{p}$$

$$\mu_2'(0) = E(x^2) = \sum_{x=0}^{\infty} x^2 q^x p = p \sum_{x=0}^{\infty} \{x(x-1) + x\} q^x$$

$$= p \sum_{x=0}^{\infty} x(x-1) q^x + p \sum_{x=0}^{\infty} x q^x$$

$$= p\{2 \cdot 1 q^2 + 3 \cdot 2 q^3 + 4 \cdot 3 q^4 + \ldots\ldots\} + \frac{q}{p}$$

$$= 2q^2 p \left\{1 + 3q + \frac{4 \cdot 3}{2!} q^2 + \frac{5 \cdot 4}{2!} q^3 \ldots + \right\} + \frac{q}{p}$$

$$= 2q^2 p (1-q)^{-3} + \frac{q}{p}$$

$$= \frac{2q^2}{p^2} + \frac{q}{p}$$

$$\therefore \quad \mu_2 = \mu_2'(0) - \bar{x}^2 = \frac{q^2}{p^2} + \frac{q}{p} = \frac{q}{p}\left(\frac{q+p}{p}\right) = \frac{q}{p^2}.$$

Ex. 10-81. *Find M.G.F. of the geometric dist.*

Sol. $$M_0(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} q^x p$$

$$= \sum_{x=0}^{\infty} (qe^t)^x p$$

$$= p \cdot \frac{1}{1-qe^t}.$$

Ex. 10-82. *If two indepen*

distribution, show that the co

Sol. Let x_1, x_2 both follo

Now $P(x_1 = r / x_1 + x$

which is discrete uniform dist

Ex. 10-83. *A population i*

affects a proportion p of the p

find the proportion of dying d

Sol. Prob. of an attack aff

$\therefore$ Prob. of an attack not

The individuals dying du

$(r-1)$ times in the first $(n-1$

Now prob. of having dise

where

$\therefore$ Prob. of dying during

which also gives the proporti

10-6. Negative Binomial dist

In the last question, prop

Ex. 10-82. *If two independent random variables x_1 and x_2 have the same geometric distribution, show that the conditional distribution of x_1 given $x_1 + x_2 = n$ is uniform.*

Sol. Let x_1, x_2 both follow the same geometric distribution

$$P(x) = pq^x, x = 0, 1, 2, \ldots$$

Now
$$P(x_1 = r / x_1 + x_2 = n) = \frac{P(x_1 = r;\, x_1 + x_2 = n)}{P(x_1 + x_2 = n)}$$

$$= \frac{P(x_1 = r;\, x_2 = n - r)}{\sum_{s=0}^{n} P(x_1 = s;\, x_2 = n - s)}$$

$$= \frac{P(x_1 = r).P(x_2 = n - r)}{\sum_{s=0}^{n} P(x_1 = s).P(x_2 = n - s)}$$

($\because x_1, x_2$ are independent)

$$= \frac{pq^r . pq^{n-r}}{\sum_{s=0}^{n} pq^s . pq^{n-s}}$$

$$= \frac{1}{n+1} \qquad r = 0, 1, 2, \ldots\ldots$$

which is discrete uniform dist.

Ex. 10-83. *A population is subjected to recurring attacks of a disease and each attack affects a proportion p of the population. Assuming that r attacks are fatal to the individual, find the proportion of dying during the nth exposure.*

Sol. Prob. of an attack affecting an individual $= p$

$\therefore$ Prob. of an attack not affecting an individual $= 1 - p$.

The individuals dying during the nth exposure will be those who have had the disease $(r-1)$ times in the first $(n-1)$ exposures and catch it again.

Now prob. of having disease $(r-1)$ times in the first $(n-1)$ exposures

$$= {}^{n-1}c_{r-1}\, p^{r-1}\, q^{n-r}$$

where $\qquad q = 1 - p$

$\therefore$ Prob. of dying during nth exposure

$$= ({}^{n-1}c_{r-1}\, p^{r-1}\, q^{n-r})(p) = {}^{n-1}c_{r-1}\, q^{n-r}\, p^r$$

which also gives the proportion of dying during the nth exposure.

10-6. Negative Binomial distribution

In the last question, proportion of individuals dying during the nth exposure

$$= {}^{n-1}c_{r-1}\, p^r\, q^{n-r}$$

Since death does not commence until the rth exposure, the proportions of death at the rth, $(r+1)$th,... exposure are

$$p^r, rqp^r, \frac{r(r+1)}{2!} q^2 p^r, \ldots$$

which are the successive terms in the binomial expansion, with negative index, of $p^r(1-q)^{-r}$

$$\therefore \qquad \sum_{n=r}^{\infty} {}^{n-1}c_{r-1}\, q^{n-r} p^r = p^r(1-q)^{-r} = 1$$

The dist. (changing n to $x + r$)

$$P(x) = {}^{x+r-1}c_{r-1}\, q^x p^r, \; x = 0, 1, 2, \ldots\ldots$$

is called Negative Binomial distribution.

Remark. Put $\quad p = \dfrac{1}{Q}, q = \dfrac{P}{Q}$

where $\quad Q - P = 1$

$$\therefore \qquad P(x) = {}^{r+x-1}c_x\, Q^{-r-x} P^x \qquad (\because\ {}^{x+r-1}c_{r-1} = {}^{x+r-1}c_x)$$

$$= {}^{-r}c_x\, Q^{-r-x}(-P)^x$$

$\Rightarrow$ All the quantities like mean, variance etc., for negative binomial variate can be written from those of binomial variate by replacing

n by $-r$; q by Q and p by $-P$

e.g., Mean $= (-r)(-P) = rP$

Variance $= (-r)(-P)Q = rPQ$ and so on.

10.6.1. Mean and Variance

$$\bar{x} = \sum_{x=0}^{\infty} x.\ {}^{x+r-1}c_{r-1}\, q^x p^r$$

$$= p^r \left\{ rq + 2\frac{(r+1)r}{2!} q^2 + \frac{3(r+2)(r+1)r}{3!} q^3 + \ldots \right\}$$

$$= rqp^r \left\{ 1 + (r+1)q + \frac{(r+2)(r+1)}{2!} q^2 \ldots \right\}$$

$$= rqp^r(1-q)^{-r-1} = \frac{rq}{p}$$

$$\mu_2'(0) = \sum_{x=0}^{\infty} x^2\ {}^{x+r-1}c_{r-1}\, q^x p^r$$

$$= \sum_{x=0}^{\infty} \{x(x-1) + x\}\ {}^{x+r-1}c_{r-1}\, q^x p^r$$

$\therefore$

Since $\dfrac{1}{p} > 1, \mu_2 > \bar{x}$.

10.6.2. Recurrence Formul

Let x be a negative binor

$P(x) = {}^{k+x-1}c_x\, q^x p^k, x$

Also

$\therefore$

Differentiating *w.r.t.* q

Now

$\therefore$

also $\quad \dfrac{d}{dq}\left(\dfrac{q}{p}\right) = \dfrac{d}{dq}\Big($

he proportions of death at the

negative index, of $p^r(1-q)^{-r}$

$= 1$

, 1, 2,.......

$(\because {}^{x+r-1}c_{r-1} = {}^{x+r-1}c_x)$

ative binomial variate can be

and so on.

$+\frac{3(r+2)(r+1)r}{3!}q^3+\dots\}$

$\frac{+2)(r+1)}{2!}q^2\dots\}$

r

${}^1c_{r-1}q^xp^r$

$$= p^r\left\{2.1\frac{(r+1)r}{2!}q^2+3\cdot2\,\frac{(r+2(r+1)r}{3!}q^3+\dots\right\}+\frac{rq}{p}$$

$$= p^r(r+1)rq^2(1-q)^{-r-2}+\frac{rq}{p}=\frac{r(r+1)q^2}{p^2}+\frac{rq}{p}$$

$$\therefore \quad \mu_2=\mu_2'(0)-\bar{x}^2$$

$$=\frac{r(r+1)q^2}{p^2}+\frac{rq}{p}-\frac{r^2q^2}{p^2}$$

$$=\frac{rq}{p^2}(q+p)=\frac{rq}{p^2}$$

Since $\frac{1}{p}>1$, $\mu_2>\bar{x}$.

10.6.2. Recurrence Formula for Moments about Mean

Let x be a negative binomial variate with parameters k and p. Then prob. f^n of x is

$P(x)={}^{k+x-1}c_x\,q^xp^k,\ x=0,1,2,\dots$

Also $\quad \bar{x}=\frac{kq}{p}.$

$$\therefore \quad \mu_r = E\left(x-\frac{kq}{p}\right)^r$$

$$=\sum_{x=0}^{\infty}\left(x-\frac{kq}{p}\right)^r\ {}^{k+x-1}c_x\,q^xp^k$$

Differentiating *w.r.t.* q

$$\frac{d\mu_r}{dq}=\sum_{x=0}^{\infty}{}^{k+x-1}c_x\left[r\left(x-\frac{kq}{p}\right)^{r-1}\left\{-k\frac{d}{dq}\left(\frac{q}{p}\right)\right\}q^xp^k\right.$$

$$\left.+\left(x-\frac{kq}{p}\right)^r\left\{xq^{x-1}p^k+q^xkp^{k-1}\frac{dp}{dq}\right\}\right]$$

Now $\quad p=1-q$

$$\therefore \quad \frac{dp}{dq}=-1$$

also $\quad \frac{d}{dq}\left(\frac{q}{p}\right)=\frac{d}{dq}\left(\frac{1}{p}-1\right)=-\frac{1}{p^2}\cdot\frac{dp}{dq}$

$$= \frac{1}{p^2}$$

$$\therefore \quad \frac{d\mu_r}{dq} = \sum_{x=0}^{\infty} {}^{k+x-1}c_x \left[-\frac{rk}{p^2}\left(x - \frac{kq}{p}\right)^{r-1} q^x p^k \right.$$

$$\left. + \left(x - \frac{kq}{p}\right)^r q^{x-1} p^k \left(x - \frac{kq}{p}\right) \right]$$

$$= \frac{-rk}{p^2}\mu_{r-1} + \frac{1}{q}\sum_{x=0}^{\infty}\left(x - \frac{kq}{p}\right)^{r+1} {}^{k+x-1}c_x\, q^x p^k$$

$$\therefore \quad \frac{d\mu_r}{dq} + \frac{rk}{p^2}\mu_{r-1} = \frac{\mu_{r+1}}{q}$$

$$\Rightarrow \quad \mu_{r+1} = q\left\{\frac{d\mu_r}{dq} + \frac{rk}{p^2}\mu_{r-1}\right\}.$$

10.6.3. Moment Generating Function and Cumulants

$$M_0(t) = \sum_{x=0}^{\infty} e^{tx} \cdot {}^{x+r-1}c_{r-1}\, q^x p^r$$

$$= p^r \sum_{x=0}^{\infty} {}^{x+r-1}c_{r-1}(qe^t)^x = p^r(1-qe^t)^{-r}$$

$\therefore$ Cumulative f^n is given by

$$K_0(t) = \log M_0(t) = r\log p - r\log(1-qe^t)$$

$$= -r\log\left\{\frac{1}{p} - \frac{q}{p}e^t\right\} = -r\log\left\{1 - \frac{q}{p}(e^t - 1)\right\}$$

$$= -r\log\left\{1 - \frac{q}{p}\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \ldots\right)\right\}$$

$$= r\left[\frac{q}{p}\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \ldots\right) + \frac{1}{2}\frac{q^2}{p^2}\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \ldots\right)^2\right.$$

$$\left. + \frac{1}{3}\frac{q^3}{p^3}\left(t + \frac{t^2}{2!}\ldots\right)^3 + \frac{1}{4}\frac{q^4}{p^4}(t + \ldots\ldots)^4 + \ldots\ldots\right]$$

$\therefore \quad k_1 = \frac{rq}{p}, \; k_2 =$

and

Ex. 10-84. *Find the lir*

but $\frac{rq}{p} = m$ *(a finite consta*

Sol. For negative binor

Let $r \to \infty, q \to 0$ so

$\therefore$

which is the probability fu

$$\frac{kq}{p}\Big)^{r-1} q^x p^k$$

$$\frac{kq}{p}\Big)\Big]$$

$$\frac{q}{p}\Big)^{r+1} {}^{k+x-1}c_x q^x p^k$$

$$= p^r(1-qe^t)^{-r}$$

$$e^t - 1)\Big\}$$

$$\Big\}$$

$$t+\frac{t^2}{2!}+\frac{t^3}{3!}+\ldots\Big)^2$$

$$^4 + \ldots\ldots\Big]$$

$$\therefore \qquad k_1 = \frac{rq}{p},\ k_2 = \frac{rq}{p^2}, k_3 = r\left\{\frac{q}{p}+\frac{3q^2}{p^2}+\frac{2q^3}{p^3}\right\}$$

$$= \frac{rq}{p^3}(1+q)$$

and

$$k_4 = r\left\{\frac{q}{p}+\frac{7q^2}{p^2}+12\frac{q^3}{p^3}+6\frac{q^4}{p^4}\right\}$$

$$= \frac{rq}{p^4}\{1+4q+q^2\}.$$

Ex. 10-84. *Find the limit of negative binomial distribution when* $r \to \infty$ *and* $q \to 0$ *but* $\frac{rq}{p} = m$ *(a finite constant).*

Sol. For negative binomial dist.

$$P(x) = {}^{x+r-1}c_{r-1}\, q^x p^r$$

$$= \frac{(x+r-1)!}{(r-1)!\,x!} q^x p^r$$

$$= \frac{(x+r-1)(x+r-2)\ldots r}{r^x . x!}\left(\frac{rq}{p}\right)^x (1-q)^{r+x}$$

$$= \frac{1}{x!}\left\{\left(1+\frac{x-1}{r}\right)\left(1+\frac{x-2}{r}\right)\ldots\left(\frac{r}{r}\right)\right\}\left(\frac{rq}{p}\right)^x$$

$$(1-q)^x . \left(1-\frac{mp}{r}\right)^r$$

$$= \frac{1}{x!}\left\{\left(1+\frac{x-1}{r}\right)\left(1+\frac{x-2}{r}\right)\ldots\frac{r}{r}\right\}\left(\frac{rq}{p}\right)^x (1-q)^x$$

$$\left\{\left(1-\frac{mp}{r}\right)^{-\frac{r}{mp}}\right\}^{-mp}$$

Let $r \to \infty, q \to 0$ so that $\frac{rq}{p} = m$. Then $p \to 1$.

$$\therefore \qquad P(x) \to \frac{m^x e^{-m}}{x!}$$

which is the probability function of Poisson dist.

10.7. Hypergeometric Distribution

Suppose an urn contains Np white and Nq blue balls $(p+q=1)$ and r balls are to be drawn one at a time without replacement. Let $P(x)$ be the prob. that out of r balls drawn x are white. Then

$$P(x) = \frac{{}^{Np}c_x \; {}^{Nq}c_{r-x}}{{}^{N}c_r}$$

$$= {}^{r}c_x \frac{(Np)^{(x)}\,(Nq)^{(r-x)}}{N^{(r)}}$$

where $x^{(r)} = x(x-1)\ldots\ldots(x-r+1)$

Consider $\quad (1+y)^{Np}(1+y)^{Nq} = \left(\sum_{s=0}^{Np} {}^{Np}c_s y^s\right)\left(\sum_{t=0}^{Nq} {}^{Nq}c_t y^t\right)$

and $\quad (1+y)^N = \sum_{r=0}^{N} {}^{N}c_r y^r$

Since $(1+y)^{Np}(1+y)^{Nq} = (1+y)^N$, equating co-efficients of y^r.

$${}^{N}c_r = \sum_{x=0}^{r} {}^{Np}c_x \; {}^{Nq}c_{r-x}$$

$$\therefore \quad \sum_{x=0}^{r} P(x) = 1$$

$\therefore$ $P(x)$ can be taken to be a probability density function. The distribution

$$P(x) = {}^{r}c_x \frac{(Np)^{(x)}(Nq)^{(r-x)}}{N^{(r)}}, \; x = 0, 1, 2, \ldots r$$

is called **Hypergeometric Distribution.**

10.7.1. Mean and Variance of Hypergeometric Distribution

$$\bar{x} = \sum_{x=0}^{r} x.\frac{{}^{Np}c_x . {}^{Nq}c_{r-x}}{{}^{N}c_r}$$

$$= \sum_{x=1}^{r} x.\frac{{}^{Np}c_x . {}^{Nq}c_{r-x}}{{}^{N}c_r} = Np\sum_{x=1}^{r} \frac{{}^{Np-1}c_{x-1} \; {}^{Nq}c_{r-x}}{{}^{N}c_r}$$

$$= \frac{Np}{{}^{N}c_r} \; {}^{Np+Nq-1}c_{r-1} = \frac{Np}{{}^{N}c_r} \; {}^{N-1}c_{r-1} = rp$$

$$\mu_2'(0) = \sum_{x=0}^{r} x^2 P(x) = \sum_{x=0}^{r} x(x-1)P(x) + \sum_{x=0}^{r} xP(x)$$

$\therefore$ $\mu_2'(0$

$\therefore$

$\therefore$

Ex. 10-85. *Find differenti*
deduce the values of moments

Sol.

Let $\quad F(\alpha, \beta;$

Then

$F(\alpha, \beta, \gamma, y)$ satisfies the

$q=1)$ and r balls are to be

. that out of r balls drawn x

$y^t\Big)$

of y^r.

The distribution

$= 0, 1, 2, \ldots r$

$\sum_{x=1}^{r} \frac{{}^{Np-1}c_{x-1}\,{}^{Nq}c_{r-x}}{{}^{N}c_r}$

${}^{N-1}c_{r-1} = rp$

$P(x) + \sum_{x=0}^{r} xP(x)$

$$= \sum_{x=0}^{r} x(x-1)\frac{{}^{Np}c_x\,{}^{Nq}c_{r-x}}{{}^{N}c_r} + rp$$

$$= \frac{(Np)(Np-1)}{{}^{N}c_r}\sum_{x=2}^{r} {}^{Np-2}c_{x-2}\,{}^{Nq}c_{r-x} + rp$$

$$= \frac{(Np)(Np-1)}{{}^{N}c_r}\cdot {}^{Np+Nq-2}c_{r-2} + rp$$

$$= \frac{(Np)(Np-1)}{{}^{N}c}\,{}^{N-2}c_{r-2} + rp$$

$$\therefore \qquad \mu_2'(0) - rp = (Np).(Np-1).\frac{{}^{N-2}c_{r-2}}{{}^{N}c_r} = (Np)(Np-1)\frac{r(r-1)}{N(N-1)}$$

$$\therefore \qquad \mu_2'(0) = \frac{rp(Np-1)(r-1)}{N-1} + rp$$

$$\therefore \qquad \mu_2 = \mu_2'(0) - \bar{x}^2 = \frac{rp(Np-1)(r-1)}{N-1} + rp - r^2p^2$$

$$= \frac{rp}{N-1}\{Npr - Np - r + 1 + N - 1 - rpN + rp\}$$

$$= \frac{rpq}{N-1}(N-r).$$

Ex. 10-85. *Find differential equation satisfied by $M_0(t)$ of hypergeometric dist. and deduce the values of moments about mean.*

Sol.
$$M_0(t) = \sum_{x=0}^{r} e^{tx}\,{}^{r}c_x \frac{(Np)^{(x)}(Nq)^{(r-x)}}{N^{(r)}}$$

$$= \sum_{x=0}^{r} e^{tx}\,{}^{r}c_x \frac{(Np)^{(x)}(Nq)^{(r)}}{N^{(r)}(Nq-r+x)^{(x)}}$$

$$= \frac{(Nq)^{(r)}}{N^{(r)}}\sum_{x=0}^{r}\frac{(Np)^{(x)}.r^{(x)}e^{tx}}{(Nq-r+x)^{(x)}x!}$$

Let
$$F(\alpha,\beta;\gamma,y) = 1 + \frac{\alpha\beta}{\gamma}\frac{y}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{y^2}{2!}\ldots$$

Then
$$M_0(t) = \frac{(Nq)^{(r)}}{N^{(r)}}F(-r,-Np;Nq-r+1,e^t)$$

$F(\alpha,\beta,\gamma,y)$ satisfies the different equation

$$y(1-y)\frac{d^2F}{dy^2}+\{\gamma-(\alpha+\beta+1)y\}\frac{dF}{dy}-\alpha\beta F=0$$

Put $\quad y = e^t$

Then diff. eq. reduces to

$$(1-e^t)\frac{d^2F}{dt^2}+\frac{dF}{dt}\{\gamma-(\alpha+\beta)e^t-1\}-\alpha\beta e^t F=0$$

$\therefore M_0(t)$ satisfies the equation

$$(1-e^t)\frac{d^2M_0(t)}{dt^2}+\frac{dM_0(t)}{dt}\{Nq-r+1-(-r-Np)e^t-1\}-rNpe^t M_0(t)=0$$

or $$(1-e^t)\left\{\frac{d^2M_0(t)}{dt^2}-(r+Np)\frac{dM_0(t)}{dt}+rNpM_0(t)\right\}+N\frac{dM_0(t)}{dt}-rNpM_0(t)=0$$

To find mean put $M_0(t)=\sum\limits_{s=0}\mu'_s\dfrac{t^s}{s!}$

Then

$$(1-e^t)\left\{\sum_{s=2}\mu'_s\frac{t^{s-2}}{(s-2)!}-(r+Np)\sum_{s=1}\mu'_s\frac{t^{s-1}}{(s-1)!}+rNp\sum_{s=0}\mu'_s\frac{t^s}{s!}\right\}$$

$$+N\sum_{s=1}\mu'_s\frac{t^{s-1}}{(s-1)!}-rNp\sum_{s=0}\mu'_s\frac{t^s}{s!}=0$$

Put $\quad t=0$

$\therefore \quad N\mu'_1-rNp=0$ or $\mu'_1=rp$

Now $\quad M_{\bar{x}}(t)=e^{-rpt}M_0(t)$

$\therefore \quad M_0(t)=e^{rpt}M_{\bar{x}}(t)$

Substituting in the diff. eq.

$$(1-e^t)\left[\frac{d^2M_{\bar{x}}(t)}{dt^2}+\{r(p-q)-Np\}\frac{dM_{\bar{x}}(t)}{dt}+(N-r)pqrM_{\bar{x}}(t)\right]+N\frac{dM_{\bar{x}}(t)}{dt}=0$$

To find moments about mean put

$$M_{\bar{x}}(t) = \sum_{s=0}\mu_s\frac{t^s}{s!} \text{ and } e^t=1+t+\frac{t^2}{2!}+\ldots\ldots$$

$$-\sum_{i=1}^{\infty}\frac{t^i}{i!}\left[\sum_{s=2}\mu_s\frac{t^{s-2}}{(s-2)!}+\{r(p-q)-Np\}\sum_{s=1}\mu_s\frac{t^{s-1}}{(s-1)!}\right.$$

$$\left.+(N-r)pqr\sum_{s=0}\mu_s\frac{t^s}{s!}\right]+N\sum_{s=1}\mu_s\frac{t^{s-1}}{(s-1)!}=0$$

Equating co-efficients of t, t^2

μ

μ

Ex. 10-86. *Show that when* N

Sol. For hypergeometric dist.

$P(x)={}^rc_x\,\{(Np)(Np-1)\ldots($

$={}^rc_x\,\{p\left(p-\frac{1}{N}\right)\ldots\left(p-\right.$

Let $\quad N\to\infty$

Then $\quad P(x)\to$

which is the probability function f

Ex. 10-87. *Deduce the mo*
hypergeometric dist.

Sol. For hypergeometric dist.

μ

and μ

Let $\quad N\to\infty$

Then $\quad \mu_2$ for I

and $\quad \mu_3$ for I

10.8. Multinomial Distribution

Let there be a series of n ind
several outcomes say $E_1, E_2, \ldots\ldots$
trial where

$-\alpha\beta F = 0$

$\alpha\beta e^t F = 0$

$^t - 1\} - rNpe^t M_0(t) = 0$

$N\dfrac{dM_0(t)}{dt} - rNpM_0(t) = 0$

$+ rNp\sum_{s=0} \mu_s' \dfrac{t^s}{s!}\Big\}$

$\mu_s' \dfrac{t^{s-1}}{(s-1)!} - rNp\sum_{s=0} \mu_s' \dfrac{t^s}{s!} = 0$

$qrM_{\bar{x}}(t)\Big] + N\dfrac{dM_{\bar{x}}(t)}{dt} = 0$

$+ t + \dfrac{t^2}{2!} + \ldots\ldots$

$\sum \mu_s \dfrac{t^{s-1}}{(s-1)!}$

$\mu_s \dfrac{t^s}{s!}\Big] + N\sum_{s=1} \mu_s \dfrac{t^{s-1}}{(s-1)!} = 0$

Equating co-efficients of $t, t^2, \ldots\ldots$

$$\mu_2 = \frac{rpq(N-r)}{N-1}$$

$$\mu_3 = \frac{rpq(q-p)(N-r)(N-2r)}{(N-1)(N-2)}.$$

Ex. 10-86. *Show that when* $N \to \infty$ *hypergeometric dist. tends to binomial dist.*

Sol. For hypergeometric dist.

$$P(x) = {}^r c_x \frac{\{(Np)(Np-1)\ldots(Np-x+1)\}\{Nq)(Nq-1)\ldots(Nq-r+x+1)\}}{N(N-1)\ldots\ldots(N-r+1)}$$

$$= {}^r c_x \frac{\left\{p\left(p-\frac{1}{N}\right)\ldots\left(p-\frac{x-1}{N}\right)\right\}\ldots\left\{q\left(q-\frac{1}{N}\right)\ldots\left(q-\frac{r-x-1}{N}\right)\right\}}{\frac{N}{N}\left(1-\frac{1}{N}\right)\ldots\left(1-\frac{r-1}{N}\right)}$$

Let $\quad N \to \infty$

Then $\quad P(x) \to {}^r c_x p^x q^{r-x}$

which is the probability function for binomial dist.

Ex. 10-87. *Deduce the moments (about mean) of binomial dist. from those of hypergeometric dist.*

Sol. For hypergeometric dist.

$$\mu_2 = \frac{rpq(N-r)}{N-1} = \frac{rpq\left(1-\frac{r}{N}\right)}{1-\frac{1}{N}}$$

and

$$\mu_3 = \frac{rpq(q-p)(N-r)(N-2r)}{(N-1)(N-2)}$$

$$= \frac{rpq(q-p)\left(1-\frac{r}{N}\right)\left(1-\frac{2r}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)}$$

Let $\quad N \to \infty$

Then $\quad \mu_2$ for B.D. $= rpq$

and $\quad \mu_3$ for B.D. $= rpq(q-p)$.

10.8. Multinomial Distribution

Let there be a series of n independent trials where each trial may result in one of the several outcomes say $E_1, E_2, \ldots\ldots E_k$ with respective probabilities $p_1, p_2, \ldots\ldots p_k$ in each trial where

$$p_1 + p_2 + \ldots\ldots + p_k = 1$$

Suppose the event E_i occurs x_i times $(i = 1, 2, ...k)$

Then $x_1 + x_2 + ...+ x_k = n$

By the theorem of compound prob., prob. of E_1 occurring x_1 times, E_2 occurring x_2 times and so on in any fixed definite order $= p_1^{x_1}\, p_2^{x_2} .....p_k^{x_k}$

Now out of n trials x_1 trials can be had in ${}^n c_{x_1}$ ways and out of remaining $(n - x_1)$ trials x_2 trials can be had in ${}^{n-x_1} c_{x_2}$ ways and so on.

$\therefore$ The total number of ways of getting $E_1 - x_1$ times, $E_2 - x_2$ times,... $E_k - x_k$ times

$$= {}^n c_{x_1} . {}^{n-x_1} c_{x_2} ... {}^{n-x_1-x_2.......-x_{k-1}} c_{x_k}$$

$$= \frac{n!}{x_1!(n-x_1)!} . \frac{(n-x_1)!}{x_2!(n-x_1-x_2)!} ... \frac{(n-x_1-x_2.....-x_{k-1})!}{x_k!(n-x_1......-x_k)!}$$

$$= \frac{n!}{x_1, x_2!......x_k!} \qquad (\because (n - x_1 ...- x_k) = 0! = 1)$$

By total prob. theorem, prob. of getting $E_1 - x_1$ times, $E_2 - x_2$ times,...... $E_k - x_k$ times

$$= \frac{n!}{x_1!x_2!.......x_k!} p_1^{x_1} p_2^{x_2} ......p_k^{x_k}$$

$$\text{Total prob.} = \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} .... \sum_{x_k=0}^{(n-x_1...x_{k-1})} \frac{n!}{x_1!x_2!.......x_k!} p_1^{x_1} p_2^{x_2} ....p_k^{x_k}$$

Consider $(p_1 + p_2......+ p_k)^n = \{p_1 + (p_2 + p_3.......+ p_k)\}^n$

$$= \sum_{x_1=0}^{n} {}^n c_{x_1} p_1^{x_1} (p_2 + p_3 ... p_k)^{n-x_1}$$

$$= \sum_{x_1=0}^{n} {}^n c_{x_1} p_1^{x_1} \left\{ \sum_{x_2=0}^{n-x_1} {}^{n-x_1} c_{x_2} p_2^{x_2} (p_3 + ...+ p_k)^{n-x_1-x_2} \right\}$$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} {}^n c_{x_1} {}^{n-x_1} c_{x_2} p_1^{x_1} p_2^{x_2} (p_3 + ...+ p_k)^{n-x_1-x_2}$$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} .... \sum_{x_k=0}^{n-x_1...-x_{k-1}} \left\{ {}^n c_{x_1} {}^{n-x_1} c_{x_2} ... {}^{n-x_1...-x_{k-1}} c_{x_k} . p_1^{x_1} p_2^{x_2} ... p_k^{x_k} \right\}$$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} .... \sum_{x_k=0}^{n-x_1...-x_{k-1}} \frac{n!}{x_1!x_2!........x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$$

$\therefore$ Total prob. $= (p_1 + p_2 + ...p_k)^n = 1$

Hence the function

$$P(x_1, x_2, ...x_k) =$$

can be taken to be probability functior

by $P(x_1, x_2, ...x_k)$ together with the v

10.8.1. Moment Generating Functi
Distribution

Now $M_0(t_1, t_2 ..., t_n) = E\{e^{t_1 x_1 + t_2 x_2}$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} ... \sum_{x_k=0}^{n-x_1...-x_{k-1}} \frac{n!}{x_1!x_2!......}$$

$$= \sum_{x_1} \sum_{x_2} ... \sum_{x_k} \frac{n!}{x_1!x_2!...x_k!} (p_1 e$$

$$= (p_1 e^{t_1} + p_2 e^{t_2} ....+ p_k e^{t_k})^n$$

$$\therefore E(x_i) = \left\{ \frac{\partial M_0}{\partial t_i} \right\}_{t_j = 0,\ j=1,2...k}$$

$$= \{n p_i e^{t_i} (p_1 e^{t_1} + p_2 e^{t_2} ...$$

$$= n p_i$$

$$E(x_i^2) =$$

$$=$$

$$=$$

$\therefore$ Va

$$E(x_i x_j) =$$

$$= \{n(n-1) p_i p_j e^{t_i} \cdot e^{t_j} (p_1 e^{t_1} + ...+$$

$$= n(n-1) p_i p_j$$

g x_1 times, E_2 occurring x_2

nd out of remaining $(n-x_1)$

$_2 - x_2$ times,... $E_k - x_k$ times

!

$(\because (n-x_1 \ldots -x_k) = 0! = 1)$

$-x_2$ times,...... $E_k - x_k$ times

$\ldots\ldots p_k^{x_k}$

n

$\ldots p_k^{x_k}\}$

Hence the function

$$P(x_1, x_2, \ldots x_k) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots\ldots p_k^{x_k}$$

can be taken to be probability function. The dist. formed.

by $P(x_1, x_2, \ldots x_k)$ together with the values of $x_1, x_2, \ldots x_k$ is called Multinomial **Dist.**

10.8.1. Moment Generating Function, Moments, Covariance, etc., for Multinomial Distribution

Now $M_0(t_1, t_2 \ldots, t_n) = E\left\{e^{t_1x_1 + t_2x_2 + \ldots\ldots + t_kx_k}\right\}$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} \cdots \sum_{x_k=0}^{n-x_1 \ldots -x_{k-1}} \frac{n!}{x_1!x_2!\ldots\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} . e^{t_1x_1 + \ldots + t_kx_k}$$

$$= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} \frac{n!}{x_1!x_2!\ldots x_k!} \left(p_1 e^{t_1}\right)^{x_1} \ldots \left(p_k e^{t_k}\right)^{x_k}$$

$$= (p_1 e^{t_1} + p_2 e^{t_2} \ldots\ldots + p_k e^{t_k})^n$$

$$\therefore E(x_i) = \left\{\frac{\partial M_0}{\partial t_i}\right\}_{t_j=0,\, j=1,2\ldots k}$$

$$= \{np_i e^{t_i} (p_1 e^{t_1} + p_2 e^{t_2} \ldots + p_k e^{t_k}\}^{n-1}\}_{\substack{t_j=0 \\ j=1,2\ldots k}}$$

$$= np_i$$

$$E(x_i^2) = \left\{\frac{\partial^2 M_0}{\partial t_i}\right\}_{\substack{t_j=0 \\ j=1,2,\ldots k}}$$

$$= \{np_i e^{t_i} (p_1 e^{t_1} + \ldots\ldots + p_k e^{t_k})^{n-1}$$
$$+ n(n-1) p_i^2 e^{2t_i} (p_1 e^{t_1} + \ldots\ldots p_k e^{t_k})^{n-2}\}_{\substack{t_j=0 \\ j=1,2,\ldots k}}$$

$$= np_i + n(n-1)p_i^2$$

$\therefore$ $$\text{Var }(x_i) = np_i + n(n-1)p_i^2 - n^2 p_i^2 = np_i(1-p_i)$$

$$E(x_i x_j) = \left\{\frac{\partial^2 M_0}{\partial t_j\, \partial t_i}\right\}_{t_1=0,\ldots\ldots t_k=0}$$

$$= \left\{n(n-1) p_i p_j e^{t_i} \cdot e^{t_j} (p_1 e^{t_1} + \ldots\ldots + p_k e^{t_k})^{n-2}\right\}_{t_1=0,\, t_2=0 \ldots t_k=0}$$

$$= n(n-1)\, p_i p_j$$

$$\therefore \text{Cov. } (x_i, x_j) = E(x_i x_j) - E(x_i)E(x_j)$$

$$= n(n-1)p_i p_j - n^2 p_i p_j = -np_i p_j$$

Aliter

$$E(x_i) = \sum x_i \frac{n!}{x_1!x_2!......x_k!} p_1^{x_1} p_2^{x_2} ......p_k^{x_k}$$

$$= p_i \frac{\partial}{\partial p_i} \sum \frac{n!}{x_1!x_2!......x_k!} p_1^{x_1} ....p_k^{x_k}$$

$$= p_i \frac{\partial}{\partial p_i} (p_1 + p_2......+p_k)^n$$

$$= np_i(p_1 + p_2...+p_k)^{n-1} = np_i$$

$$E(x_i^2) = \sum x_i^2 \frac{n!}{x_1!x_2!......x_k!} p_1^{x_1} p_2^{x_2} ......p_k^{x_k}$$

$$= p_i \frac{\partial}{\partial p_i} \sum x_i \frac{n!}{x_1!x_2!.......x_k!} p_1^{x_1} p_2^{x_2} ....p_k^{x_k}$$

$$= p_i \frac{\partial}{\partial p_i} \left\{ p_i \frac{\partial}{\partial p_i} \sum x_i \frac{n!}{x_1!x_2!.......x_k!} p_1^{x_1} p_2^{x_2} ....p_k^{x_k} \right\}$$

$$= p_i \frac{\partial}{\partial p_i} \left\{ p_i \frac{\partial}{\partial p_i} (p_1 + p_2 + ... + p_k)^n \right\}$$

$$= p_i \frac{\partial}{\partial p_i} \{np_i (p_1 + p_2...+p_k)^{n-1}\}$$

$$= p_i \{n(p_1 + p_2......+p_k)^{n-1} + n(n-1)p_i(p_1 + ...+p_k)^{n-2}\}$$

$$= p_i \{n + n(n-1)p_i\} = np_i + n(n-1)p_i^2.$$

$$E(x_i x_j) = \sum x_i x_j \frac{n!}{x_1!x_2!....x_k!} p_1^{x_1} p_2^{x_2} ....p_k^{x_k}$$

$$= p_i \frac{\partial}{\partial p_i} \left\{ \sum x_j \frac{n!}{x_1!x_2!....x_k!} p_1^{x_1} p_2^{x_2} ......p_k^{x_k} \right\}$$

$$= p_i \frac{\partial}{\partial p_i} \left\{ p_j \frac{\partial}{\partial p_j} (p_1 + p_2......p_k)^n \right\}$$

$$= n(n-1)p_i p_j.$$

□□□

Bivari

11.1. Discrete Bivariate Distribut

In the case of discrete bivari and the pairs of values of X and Y for the pairs.

Let $x_1, x_2 ... x_m$ and $y_1, y_2 ... y_r$ for the pair (x_i, y_j) be denoted by

Then $\sum_{i=1}^{m} \sum_{j=1}^{m} f_{i,}$

is the total frequency.

If X and Y are random variates then the function p s.t.

$p(x_i, y_j$

is called the joint probability funct

Definitions

(1) Let $p_X(x_i$

and $p_Y(y_j$

p_X (or p_Y) is called **margin**

(2) $p_{X/Y}(x_i / y_j$

is called **conditional probabi**

Similarly $p_{Y/X}(y_j / x_i$

is called **conditional probability**

(3) X and Y are said to be ind

p

Otherwise they are said to be

$\ldots\ldots p_k^{x}$

$\ldots\ldots p_k^{x_k}$

$\ldots\ldots p_k^{x_k}$

$p_1^{x_1} p_2^{x_2} \ldots\ldots p_k^{x_k}$

$\frac{n!}{\ldots\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots\ldots p_k^{x_k} \Big\}$

$p_k)^n \Big\}$

${}^1\}$

$(n-1)p_i(p_1+\ldots+p_k)^{n-2}\}$

$-1)p_i^2.$

$\ldots p_k^{x_k}$

$p_1^{x_1} p_2^{x_2} \ldots\ldots p_k^{x_k} \Big\}$

$_k)^n \Big\}$

□□□

11

Bivariate Distribution

11.1. Discrete Bivariate Distributions

In the case of discrete bivariate distribution there are two discrete variates X, Y and the pairs of values of X and Y are considered. The frequencies or probabilities are for the pairs.

Let $x_1, x_2 \ldots x_m$ and $y_1, y_2 \ldots y_n$ be the values of X and Y respectively and the frequency for the pair (x_i, y_j) be denoted by f_{ij}.

Then $$\sum_{i=1}^{m} \sum_{j=1}^{m} f_{ij} = N \qquad \text{(say)}$$

is the total frequency.

If X and Y are random variates, and probability for the pair (x_i, y_j) is denoted by p_{ij} then the function p $s.t.$

$$p(x_i, y_j) = p_{ij}$$

is called the joint probability function of X and Y.

Definitions

(1) Let $$p_X(x_i) = p_{i1} + p_{i2} + \ldots + p_{in}$$

and $$p_Y(y_j) = p_{1j} + p_{2j} + \ldots + p_{mj}$$

p_X (or p_Y) is called **marginal probability function of X (or Y)**

(2) $$p_{X/Y}(x_i / y_j) = \frac{p_{ij}}{p_Y(y_j)}$$

is called **conditional probability function of X given $Y = y_j$**

Similarly $$p_{Y/X}(y_j / x_i) = \frac{p_{ij}}{p_X(x_i)}$$

is called **conditional probability function of Y given $X = x_i$**

(3) X and Y are said to be independent if

$$p_{ij} = p_X(x_i) \times p_Y(y_j)$$

Otherwise they are said to be dependent.

Joint discrete density function

The joint discrete density function $f_{X,Y}(\cdot,\cdot)$ is defined by

$$f_{X,Y}(x,y) = \begin{cases} P\{X = x, Y = y\} & \text{for a value pair } (x,y) \text{ of } (X,Y) \\ 0 & \text{(otherwise)} \end{cases}$$

Marginal discrete density functions

Marginal discrete density functions are defined by

$$f_X(x_k) = \sum_J f_{X,Y}(x_k, y_j)$$

$$f_Y(y_k) = \sum_i f_{X,Y}(x_i, y_k)$$

Joint Cumulative Distribution Function :

Def : Let X, Y be two random variables both defined on the same probability space. The joint cumulative distribution function of X, Y is denoted by $F_{X,Y}(\cdot,\cdot)$ and is defined as $F_{X,Y}(x,y) = P(X \le x; Y \le y)$ for all value pairs (x, y).

Properties of Cumulative Distribution Function

c.d.f. $F_{X,Y}(\cdot,\cdot)$ satisfies following properties :

(*I*)
$$\lim_{x \to -\infty} F_{X,Y}(x;y) = F_{X,Y}(-\infty, y) = 0 \text{ for all } y$$

$$\lim_{y \to -\infty} F_{X,Y}(x,y) = F_{X,Y}(x, -\infty) = 0 \text{ for all } x$$

$$\lim_{\substack{x \to \infty \\ y \to \infty}} F_{X,Y}(x,y) = F_{X,Y}(\infty, \infty) = 1$$

(*II*)
$$0 \le F_{X,Y}(x,y) \le 1$$

(*III*) $F_{X,Y}(x,y)$ is monotonically non-decreasing

(*i*) $F_{X,Y}(x_1, y) \ge F_{X,Y}(x_2, y)$ if $x_1 \ge x_2$

(*ii*) $F_{X,Y}(x, y_1) \ge F_{X,Y}(x, y_2)$ if $y_1 \ge y_2$

To prove (*i*) we observe that

$$P(X \le x_1, Y \le y) = P(X \le x_2, Y \le y) + P(x_2 < X \le x_1, Y \le y)$$

$$\Rightarrow \quad F_{X,Y}(x_1, y) = F_{X,Y}(x_2, y) + P(x_2 < X \le x_1, Y \le y)$$

$$\Rightarrow \quad F_{X,Y}(x_1, y) \ge F_{X,Y}(x_2, y)$$

Similarly (*ii*) can be proved.

(*IV*) If $x_1 < x_2$ and $y_1 < y_2$, then

$$P\{x_1 < X \le x_2, y_1 < Y \le y_2)$$

$= F_{X,}$

$-F_{}$

To prove this define

$A_1 = (X \le x_1), A_2 = (X$

$B_1 = (Y \le y_1), B_2 = (Y$

Then L.H.S. $= P\{($

$= P\{$

$= P\{$

$= P\{$

$+P$

$= F_X$

Marginal Cumulative Distri

Marginal cumulative distr

$F_X(x) = F_X$

$F_Y(y) = F_X$

Result : $F_X(x) + F_Y(y) -$

Proof : Define events A a

Then $P(A) = P($

$P(B) = P($

By additive law

$P(A \cup B) = P($

$\therefore \quad P(A \cap B) \ge P($

$\Rightarrow F_X$

$A \cap$

$A \cap$

$\therefore F_X(x)$

Sometimes, same small l

Ex. 11-1. *x and y are tw*

$f(x,y) = \frac{1}{2}$

where x and y can assume on of y for given x.

efined by

$= y\}$ for a value pair (x, y) of (X, Y)

herwise)

by

$y_j)$

$y_k)$

ied on the same probability space. The noted by $F_{XY}(\cdot,\cdot)$ and is defined as).

0 for all y

0 for all x

1

g

$y) + P(x_2 < X \le x_1, Y \le y)$

$P(x_2 < X \le x_1, Y \le y)$

$\le y_2)$

$$= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1)$$
$$-F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1)$$

To prove this define

$$A_1 = (X \le x_1),\ A_2 = (X \le x_2)$$
$$B_1 = (Y \le y_1),\ B_2 = (Y \le y_2)$$

Then

$$\begin{aligned}
\text{L.H.S.} &= P\{(A_2 - A_1) \cap (B_2 - B_1)\} \\
&= P\{A_2 \cap (B_2 - B_1)\} - P\{(A_1 \cap (B_2 - B_1)\} \\
&= P\{A_2 \cap B_2\} - P(A_2 \cap B_1) - P(A_1 \cap B_2) + P(A_1 \cap B_1) \\
&= P\{X \le x_2, Y \le y_2) - P(X \le x_2, Y \le y_1) - P(X \le x_1; Y \le y_2) \\
&\quad + P(X \le x_1; Y \le y_1) \\
&= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2 y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1)
\end{aligned}$$

Marginal Cumulative Distribution Function

Marginal cumulative distribution functions of X, Y are defined by

$$F_X(x) = F_{X,Y}(x, \infty)$$
$$F_Y(y) = F_{X,Y}(\infty, y)$$

Result : $F_X(x) + F_Y(y) - 1 \le F_{X,Y}(x, y) \le \sqrt{F_X(x) F_Y(y)}$ for all x, y.

Proof : Define events A and B such that

$$A : X \le x \text{ and } B : Y \le y.$$

Then

$$P(A) = P(X \le x) = F_X(x)$$
$$P(B) = P(Y \le y) = F_Y(y)$$

By additive law

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \le 1$$
$$\therefore \quad P(A \cap B) \ge P(A) + P(B) - 1$$
$$\Rightarrow F_X(x) + F_Y(y) - 1 \le P(A \cap B) = F_{X,Y}(x, y)$$
$$A \cap B \subseteq A \Rightarrow P(A \cap B) \le P(A) = F_X(x)$$
$$A \cap B \subseteq B \Rightarrow P(A \cap B) \le P(B) = F_Y(y)$$
$$\therefore \{P(A \cap B)\}^2 \le F_X(x)\, F_Y(y)$$
$$\Rightarrow P(A \cap B) \le \sqrt{F_X(x)\, F_Y(y)}$$
$$\therefore F_X(x) + F_Y(y) - 1 \le F_{X,Y}(x, y) \le \sqrt{F_X(x) F_Y(y)}.$$

Sometimes, same small letters are used to denote variates as well as their values.

Ex. 11-1. *x and y are two random variables having the joint density function*

$$f(x, y) = \frac{1}{27}(x + 2y),$$

where x and y can assume only the integer values 0, 1, 2. Find the conditional distribution of y for given x.

Sol. By given

$f(x, y) = \frac{1}{27}(x+2y)$, $x = 0, 1, 2$; $y = 0, 1, 2$. The table below gives various values of f

$y \rightarrow$ / $x \downarrow$	0	1	2	f_x
0	0	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{6}{27}$
1	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{5}{27}$	$\frac{9}{27}$
2	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{6}{27}$	$\frac{12}{27}$

The last column headed f_x gives the marginal probability function of x.

The table giving the values of conditional probability function of y for given x.

$y \rightarrow$ / $x \downarrow$	0	1	2
0	0	$\frac{1}{3}$	$\frac{2}{3}$
1	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$
2	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{6}{12}$

This is obtained by dividing each row by corresponding entry under f_x.

Ex. 11-2. *Two unbiased dice are tossed simultaneously. If x and y be the numbers on two dice respectively, find*

(*i*) $P\{x+y=6 \mid y=2\}$. (*ii*) $P(x-y=2)$.

Sol. Both x and y can take values 1, 2, 3, 4, 5, 6 each with probability $\frac{1}{6}$.

The joint probability function of x and y is given by

$$p(x, y) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \ \forall x \text{ and } \forall y.$$

The table listing the values of p is

$y \rightarrow$ / $x \downarrow$	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

(*Contd.*)

4	$\frac{1}{36}$	$\frac{1}{3}$
5	$\frac{1}{36}$	$\frac{1}{3}$
6	$\frac{1}{36}$	$\frac{1}{3}$
p_y	$\frac{1}{6}$	$\frac{1}{6}$

The last row gives the probabi

(*i*) $\therefore \quad P(x+y = 6 \mid y=2$

$= \frac{p\{x=}{p_y(}$

$= \frac{1}{36} \Big/ \frac{1}{6}$

(*ii*) $P(x-y=2) = P(x=$

$+P(x=$

$= \frac{1}{36} + $

Ex. 11-3. *The joint probabilit by the following table :*

$x \rightarrow$

$y \downarrow$

1

2

Find (*i*) the marginal distribu

(*ii*) the conditional dist

(*iii*) $P\{x+y \leq 3\}$.

Sol. The given table is

$x \rightarrow$ / $y \downarrow$	1
1	0·1
2	0·2
p_x	0·3

(*i*) The marginal distributio

distribution of x is given by row h

(*ii*) The conditional distribu

corresponding entry in column he

e below gives various values of f

2	f_x
$\frac{4}{27}$	$\frac{6}{27}$
$\frac{5}{27}$	$\frac{9}{27}$
$\frac{6}{27}$	$\frac{12}{27}$

bility function of x.

function of y for given x.

ig entry under f_x.

ly. If x and y be the numbers on

vith probability $\frac{1}{6}$.

y.

5	6
$\frac{1}{36}$	$\frac{1}{36}$
$\frac{1}{36}$	$\frac{1}{36}$
$\frac{1}{36}$	$\frac{1}{36}$

(*Contd.*)

4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
p_y	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The last row gives the probability function of y

(*i*) $\therefore$
$$P(x+y = 6|y=2\} = P\{x=4|y=2\}$$
$$= \frac{p\{x=4, y=2\}}{p_y\,(y=2)}$$
$$= \frac{1}{36}\Big/\frac{1}{6} = \frac{1}{6}.$$

(*ii*)
$$P(x-y=2) = P(x=3, y=1) + P(x=4, y=2) + P(x=5, y=3) + P(x=6, y=4)$$
$$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{9}.$$

Ex. 11-3. *The joint probability distribution of a pair (x, y) of random variables is given by the following table :*

$x \rightarrow$ / $y \downarrow$	1	2	3
1	0·1	0·1	0·2
2	0·2	0·3	0·1

Find (*i*) the marginal distributions

(*ii*) the conditional distribution of x given $y = 1$

(*iii*) $P\{x+y \le 3\}$.

Sol. The given table is

$x \rightarrow$ / $y \downarrow$	1	2	3	p_y
1	0·1	0·1	0·2	0·4
2	0·2	0·3	0·1	0·6
p_x	0·3	0·4	0·3	

(*i*) The marginal distribution of y is given by column headed p_y and the marginal distribution of x is given by row headed p_x.

(*ii*) The conditional distribution of x for $y = 1$ is given by dividing first row by corresponding entry in column headed p_y.

$\therefore$ Conditional distribution of x for $y = 1$ is

$$p(x / y = 1) \; : \; \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{2}$$

(iii)
$$\begin{aligned} P(x+y \le 3) &= P(x+y=2) + P(x+y=3) \\ &= P(x=1, y=1) + P(x=1, y=2) + P(x=2, y=1) \\ &= 0{\cdot}1 + 0{\cdot}2 + 0{\cdot}1 \\ &= 0{\cdot}4. \end{aligned}$$

Ex. 11-4. *Two discrete random variables X and Y have*

$$p(0,0) = \frac{2}{9},\; p(0,1) = \frac{1}{9},\; p(1,0) = \frac{1}{9},\; p(1,1) = \frac{5}{9}.$$

Test whether X and Y are independent.

Sol. The given data can be put in the form of table below

$X \rightarrow$ / $Y \downarrow$	0	1	p_Y
0	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{3}{9}$
1	$\frac{1}{9}$	$\frac{5}{9}$	$\frac{6}{9}$
p_x	$\frac{3}{9}$	$\frac{6}{9}$	

Since $p(x., y) \neq p_X(x).p_Y(y)$, X and Y are not independent.

Ex. 11-5. *Two tetrahedron (regular four-sided polyhedron each with sides labelled 1 to 4) are thrown together. Let x denote the number on the downturned face of the first tetrahedron and y the larger of the downturned numbers. Find the joint discrete density function of x and y. Also find their joint cumulative distribution function.*

Sol. Possible value pairs for (x, y) are :

(1, 1), (1, 2), (1, 3), (1, 4)
(2, 2), (2, 3), (2, 4)
(3, 3), (3, 4)
(4, 4)

Consider pair (2, 2).

Here no. 2 appears on first tetrahedron and on second tetrahedron nos. 1 and 2 both can appear.

$$\therefore \text{ prob. of pair } (2, 2) = \frac{1}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{1}{4} = \frac{2}{16}.$$

Similarly for pair (3, 3), no. 3 appear on first tetrahedron and nos. 1,2, 3 can appear on second one.

$$\therefore \text{ prob. for pair } (3, 3) = \frac{3}{16}$$

and prob. for pair (4, 4) $= \dfrac{4}{16}$

Prob. for each of the othe

$\therefore$ Joint discrete density f

(x, y) : (1, 1) (1,

$f_{x,y}(x, y)$: $\frac{1}{16}$ 1/1(

(x, y) : (3, 3) (3,

$f_{x,y}(x, y)$: $\frac{3}{16}$ 1/1(

or in another tabular forn

$x \rightarrow$ / $y \downarrow$	1
1	1/1(
2	1/1
3	1/1
4	1/1
$f_x(x)$	4/

Table for joint distributic

$x \rightarrow$ / $y \downarrow$	$x < 1$
$y < 1$	0
$1 \le y < 2$	0
$2 \le y < 3$	0
$3 \le y < 4$	0
$4 \le y$	0

This is because

$\frac{1}{2}$

$P(x+y=3)$

$+ P(x=1, y=2) + P(x=2, y=1)$

have

$p(1,0) = \frac{1}{9}$, $p(1,1) = \frac{5}{9}$.

: below

	p_Y
	$\frac{3}{9}$
	$\frac{6}{9}$

ependent.

hedron each with sides labelled 1 to
vnturned face of the first tetrahedron
nt discrete density function of x and

, 4)

d tetrahedron nos. 1 and 2 both can

dron and nos. 1,2, 3 can appear on

Prob. for each of the other pairs is $\frac{1}{16}$.

$\therefore$ Joint discrete density function is :

(x, y)	:	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(2, 2)	(2, 3)	(2, 4)
$f_{x,y}(x, y)$	:	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
(x, y)	:	(3, 3)	(3, 4)	(4, 4)				
$f_{x,y}(x, y)$	:	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$				

or in another tabular form is :

$y \downarrow$ \ $x \rightarrow$	1	2	3	4	$f_y(y)$
1	$\frac{1}{16}$				$\frac{1}{16}$
2	$\frac{1}{16}$	$\frac{2}{16}$			$\frac{3}{16}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$		$\frac{5}{16}$
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{7}{16}$
$f_x(x)$	4/16	4/16	4/16	4/16	

Table for joint distribution function is :

$y \downarrow$ \ $x \rightarrow$	$x < 1$	$1 \le x < 2$	$2 \le x < 3$	$3 \le x < 4$	$4 \le x$
$y < 1$	0	0	0	0	0
$1 \le y < 2$	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
$2 \le y < 3$	0	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$
$3 \le y < 4$	0	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{9}{16}$
$4 \le y$	0	$\frac{4}{16}$	$\frac{8}{16}$	$\frac{12}{16}$	1

This is because

$$F(1,1) = P(x \le 1, y \le 1)$$

$$= P(x=1, y=1) = \frac{1}{16}$$

$$
\begin{aligned}
F(2,3) &= P(x \le 2, y \le 3) \\
&= P(x=1, y=1) + P(x=1, y=2) + P(x=1, y=3) \\
&\quad + P(x=2, y=1) + P(x=2, y=2) + P(x=2, y=3) \\
&= \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + 0 + \frac{2}{16} + \frac{1}{16} = \frac{6}{16} \\
&\quad \{\because P(x=2, y=1) = 0 \text{ as } y > x\}
\end{aligned}
$$

similarly others.

Remark. Marginal densities can be obtained from the joint densities as shown in first table of last example. But converse is not true.

Ex. 11-6. *Let x, y be jointly discrete random variables such that each x and y have at most two values. Prove or disprove : x and y are independent iff they are uncorrelated.*

Sol. Let each of the variates x, y take values 0 and 1.

$$
\begin{aligned}
&\text{let} && \ell_{x,y} = 0 \\
&\Rightarrow && \text{Cov}\,(x, y) = 0 \\
&\Rightarrow && E(xy) = E(x)\,E(y) \\
&\Rightarrow && 0.0.\,P(x=0 \cap y=0) + 1.0\,P(x=1 \cap y=0) \\
& && +0.1.\,P(x=0 \cap y=1) + 1.1\,P(x=1 \cap y=1) \\
& && = \{1.P(x=1) + 0.P(x=0)\}.\ \{1.P(y=1) + 2.P(y=0)\} \\
&\Rightarrow && P(x=1 \cap y=1) = P(x=1).P(y=1)
\end{aligned}
$$

$\Rightarrow$ x and y are independent.

Converse. Let x and y by independent

$$
\begin{aligned}
\text{Then} \quad \text{Cov}\,(x, y) &= E(xy) - E(x)E(y) \\
&= E(x).E(y) - E(x)E(y) \\
&= 0 \\
\therefore \ell_{xy} &= 0.
\end{aligned}
$$

11.2. Continuous Bivariate Distributions

In the case of continuous bivariate distribution, variates X and Y are continuous. Here various terms are defined as below :

(1) Probability Density Function. A continuous function $f_{X,Y}(\cdot,\cdot)$ *s.t.* the probability of the value of the variate to lie in infinitestimal intervals $\left[x - \frac{dx}{2}, x + \frac{dx}{2}\right]$ and $\left[y - \frac{dy}{2}, y + \frac{dy}{2}\right]$ can be expressed in the form $f_{X,Y}(x, y)dx.dy$, is called probability density function or simply the density function.

The density function $f_{X,Y}(\cdot,\cdot)$ has the following properties :

(*i*) $f_{X,Y}(x, y) \ge 0, \forall x$ and $\forall y$.

(*ii*) $\iint f_{X,Y}(x, y)\, dxdy = 1$

where the integral is extended ove

(2) Probability Differential

$f_{X,Y}(x,y)\,dx\,dy$ is called pro

$P[a \le x \le b, c \le y \le d] = \int_{x=a}^{b}$

(3) Marginal Distributions

Let $f_X($

and f_Y

Then, $f_X(x)$ [or $f_Y(y)$] is ca

$f_X(x)dx$ is called marginal c

of Y.

(4) Conditional Density Fu

$f_{X/Y}(x$

is called conditional density fun

Similarly, $f_{Y/X}(y$

is called conditional density fun

(5) Two variates X and Y ar

$f_{X,Y}($

(6) Joint Probability Distr

It is denoted by $F_{X,Y}(x, y)$

$F_{X,Y}($

It is also called **cumulat** properties :

(*i*) $F_{X,Y}(-$

(*ii* $F_{X,Y}(\infty$

(*iii*) $\frac{\partial^2}{\partial}$

(*iv*) $F_{X,Y}(x, y)$ is monoto

$+ P(x=1, y=3)$

$2) + P(x=2, y=3)$

$\frac{6}{16}$

joint densities as shown in first

such that each x and y have at
ent iff they are uncorrelated.

$0)$

$1)$

$- 2. P(y=0)\}$

$y)$

s X and Y are continuous. Here

on $f_{X,Y}(\cdot,\cdot)\, s.t.$ the probability

tervals $\left[x-\frac{dx}{2}, x+\frac{dx}{2}\right]$ and

dy, is called probability density

ties :

where the integral is extended over the entire range of (x, y). p.d.f is also denoted by $f(\cdot,\cdot)$

(2) Probability Differential

$f_{X,Y}(x, y)\,dx\,dy$ is called probability differential. Moreover

$$P[a \le x \le b, c \le y \le d] = \int_{x=a}^{b} \int_{y=c}^{d} f_{X,Y}(x, y)\,dx\,dy$$

(3) Marginal Distributions and Marginal Density Functions

Let
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$$

and
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx.$$

Then, $f_X(x)$ [or $f_Y(y)$] is called marginal function X (or Y).

$f_X(x)dx$ is called marginal distribution of X and $f_Y(y)dy$ is called marginal distribution of Y.

(4) Conditional Density Function

$$f_{X/Y}(x/y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

is called conditional density function of X given Y

Similarly,
$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

is called conditional density function of Y given X.

(5) Two variates X and Y are said to be independent (or stochastically independent) if

$$f_{X,Y}(x, y) = f_X(x).f_Y(y).$$

(6) Joint Probability Distribution Function

It is denoted by $F_{X,Y}(x, y)$ and is given by

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x, y)\,dx\,dy$$

It is also called **cumulative distribution function** and possesses the following properties :

(i) $F_{X,Y}(-\infty, y) = 0 = F_{X,Y}(x, -\infty)$

(ii $F_{X,Y}(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = 1$

(iii) $\frac{\partial^2 F_{X,Y}}{\partial x \partial y} = f_{X,Y}(x, y)$

(iv) $F_{X,Y}(x, y)$ is monotonic non-decreasing function.

Remark : The conditional density functions

$$f_{Y/X}(y/x) \qquad (\text{and } f_{X/Y}(x/y)$$

are undefined for $\quad f_X(x) = 0 \quad (\text{and } f_Y(y) = 0$

Conditional cumulative distribution function

These functions are defined as follows :

$$F_{Y/X}(y/x) = \int_{-\infty}^{y} f_{Y/X}(y/x)\, dy$$

for all x s.t $f_X(x) > 0$.

and
$$F_{X/Y}(x/y) = \int_{-\infty}^{x} f_{X/Y}(x/y)\, dx$$

for all y s.t $f_Y(y) > 0$.

Remark : Obviously $f_{Y/X}(y/x)$ is non-negative and

$$\int_{-\infty}^{\infty} f_{Y/X}(y/x)\, dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}\, dy$$

$$= \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

$$= \frac{f_X(x)}{f_X(x)} = 1.$$

Ex. 11-7. *Let $f_X(x)$ and $f_Y(y)$ be two probability density functions with corresponding cumulative distribution functions $F_X(x)$ and $F_Y(y)$ respectively. Define*

$$f_{X,Y}(x,y;\alpha) = f_X(x)\, f_Y(y)\{1 + \alpha[2F_X(x) - 1][2F_Y(y) - 1]\}, \quad -1 \le \alpha \le 1$$

Show that : (*i*) $f_{X,Y}(x,y;\alpha)$ is a joint probability density function

(*ii*) the marginals of $f_{X,Y}(x,y;\alpha)$ are $f_X(x)$ and $f_Y(y)$ respectively.

Sol. We have

$$0 \le F_X(x) \le 1$$

$$\Rightarrow \quad 0 \le 2\,F_X(x) \le 2$$

$$\Rightarrow \quad -1 \le 2\,F_X(x) - 1 \le 1$$

Similarly $\quad -1 \le 2\,F_Y(x) - 1 \le 1$

also $\quad -1 \le \alpha \le 1$

$$\therefore -1 \le \alpha\,\{2F_X(x) - 1\}\,\{2F_Y(y) - 1\} \le 1$$

$$\Rightarrow \quad 0 \le 1 + \alpha\{2F_X(x) - 1\}\,\{2F_Y(y) - 1\} \le 2$$

$\Rightarrow f_{X,Y}(x,y;\alpha) \ge 0$ as $f_X(x)$;

Also $\quad I = \int_{-\infty}^{\infty} f_{X,Y}(x.$

$= \int_{-\infty}^{\infty} f_X(x)\,.$

$= f_X(x) \int_{-\infty}^{\infty}$

Let $\quad I_1 = \int_{-\infty}^{\infty} f_Y(y)\{$

Put $F_Y($

$d\,F$

$\Rightarrow f_Y($

$= \int_0^1 (2U -$

$\therefore \quad I = f_X(x)$

Similarly $\int_{-\infty}^{\infty} f_{X,Y}(x,y;\alpha)\, dx =$

$\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y;\alpha)\, dx$

$\therefore \quad f_{X,Y}(x,y;\alpha)$ is a jo

Ex. 11-8. *For the bivariate dis*

$f(x,y)$

find $F_{X,Y}(x,y)$; $f_X(x)$, $f_Y(y)$, $f_{Y/X}($

Sol. K is given by

$1 = \int_0^1\int_0^1 K\,($

$= K\int_0^1 dx$

inctions with corresponding

y. Define

$-1]\}$, $-1 \le \alpha \le 1$

ınction

ınd $f_Y(y)$ respectively.

$\Rightarrow \; f_{X,Y}(x, y; \alpha) \ge 0$ as $f_X(x)$; $f_Y(y)$ are non-negative.

Also
$$I = \int_{-\infty}^{\infty} f_{X,Y}(x, y, \alpha) dy$$

$$= \int_{-\infty}^{\infty} f_X(x)\, f_Y(y)\{1 + \alpha[2F_x(x) - 1\}\,\{2F_Y(y) - 1\} dy$$

$$= f_X(x)\int_{-\infty}^{\infty} f_Y(y) dy + \alpha f_x(x)\{2F_x(x) - 1\} \int_{-\infty}^{\infty} f_Y(y)\{2F_Y(y) - 1\} dy$$

Let
$$I_1 = \int_{-\infty}^{\infty} f_Y(y)\{2F_Y(y) - 1\} dy$$

Put $F_Y(y) = U$

$$d\,F_Y(y) = dU$$

$$\Rightarrow \; f_Y(y) dy = dU$$

$$= \int_0^1 (2U - 1) dU = \left\{U^2 - U\right\}_0^1 = 0$$

$\therefore$
$$I = f_X(x) \qquad (\because \int_{-\infty}^{\infty} f_Y(y)\, dy = 1$$

Similarly
$$\int_{-\infty}^{\infty} f_{X,Y}(x, y; \alpha)\, dx = f_Y(y)$$

$$\therefore \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y; \alpha)\, dx dy = \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

$\therefore$ $f_{X,Y}(x, y; \alpha)$ is a joint p.d.f with marginals $f_X(x)$ and $f_Y(y)$ respectively.

Ex. 11-8. *For the bivariate distribution given by*

$$f(x, y) = K(x + y)\, I_{(0,1)}(x)\, I_{(0,1)}(y)$$

find $F_{X,Y}(x, y)$; $f_X(x)$, $f_Y(y)$, $f_{Y/X}(y / x)$, $f_{X/Y}(x / y)$, $f_{Y/X}(y / x)$.

Sol. K is given by

$$1 = \int_0^1\int_0^1 K\,(x + y)\, dx\, dy$$

$$= K\int_0^1 dx\left\{x + \frac{1}{2}\right\}$$

$$= K\left\{\frac{1}{2}+\frac{1}{2}\right\} = K$$

$$F_{X,Y}(x,y) = \left\{\int_0^x\int_0^y (x+y)\,dx\,dy\right\} I_{(0,1)}(x)\, I_{(0,1)}(y)$$

$$+\left\{\int_0^1 dy\int_0^x (x+y)\,dx\,dy\right\} I_{(0,1)}(x)\, I_{(1,\infty)}(y)$$

$$+\left\{\int_0^1 dx\int_0^y (x+y)\,dy\right\} I_{(1,\infty)}(x)\, I_{(0,1)}(y)$$

$$+ I_{(1,\infty)}(x)\, I_{(1,\infty)}(y)$$

$$= \frac{1}{2}\Big[(x^2y+xy^2)\, I_{(0,1)}(x)\, I_{(0,1)}(y) + (x^2+x)\, I_{(0,1)}(x)\, I_{(1,\infty)}(y)$$

$$+(y+y^2)\, I_{(1,\infty)}(x)\, I_{(0,1)}(y) + I_{(1,\infty)}(x)\, I_{(1,\infty)}(y)\Big]$$

$$f_X(x) = \left\{\int_0^1 (x+y)\,dy\right\} I_{(0,1)}(x) = \left(x+\frac{1}{2}\right) I_{(0,1)}(x)$$

$$f_Y(y) = \left\{\int_0^1 (x+y)\,dx\right\} I_{(0,1)}(y) = \left(y+\frac{1}{2}\right) I_{(0,1)}(y)$$

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \frac{(x+y)\, I_{(0,1)}(x)\, I_{(0,1)}(y)}{\left(x+\frac{1}{2}\right) I_{(0,1)}(x)}$$

$$= \frac{x+y}{x+\frac{1}{2}}\, I_{(0,1)}(y)$$

Similarly

$$f_{X/Y}(x/y) = \frac{x+y}{y+\frac{1}{2}}\, I_{(0,1)}(x)$$

$$F_{Y/X}(y/x) = \int_0^y \frac{x+y}{x+\frac{1}{2}}\,dy = \frac{1}{x+\frac{1}{2}}\left\{xy+\frac{y^2}{2}\right\} \qquad \text{(for } 0<y<1\text{).}$$

Ex. 11-9. *If* $F(\cdot)$ *is a cumulativ*

(*i*) *Is* $F(x,y) = F(x)+F(y)$ *a*

(*ii*) *Is* $F(x,y) = F(x)F(y)$ *a joi*

Sol. (*i*) Is not true because

$F(\infty,\infty)$

(*ii*) It is true because

(*a*) $F(\infty,\infty) = F(\infty)F(\infty) =$

(*b*) $F(-\infty,-y) = F(-\infty)F(-$

$F(x,-\infty) = F(x)F(-\infty)$

(*c*) $\frac{\partial^2 F}{\partial x\,\partial y} = \frac{\partial F(x)}{\partial x}\frac{\partial F(y)}{\partial y}$

$= f(x)\,g(y)$

(*d*) $F(x,y)$ is monotonic n

Ex. 11-10. *If x and y are two ra*

$f(x,y) = \frac{1}{9}($

Find (*i*) $p\{x\le 1, y<3\}$

(*ii*) $p\{x+y<3\}$

(*iii*) $p\{x<1/y<3\}$.

Solution.

(*i*) $P\{x\le 1, y<3\}$

(*ii*) $p(x+y<3)$

$I_{(0,1)}(y)$

$(x)\, I_{(1,\infty)}(y)$

$x)\, I_{(0,1)}(y)$

$) + (x^2 + x)\, I_{(0,1)}(x)\, I_{(1,\infty)}(y)$

$_{(1,\infty)}(x)\, I_{(1,\infty)}(y)\big]$

$\left.\frac{1}{2}\right)I_{(0,1)}(x)$

$\left.\frac{1}{2}\right)I_{(0,1)}(y)$

(for $0 < y < 1$).

Ex. 11-9. *If $F(\cdot)$ is a cumulative distribution f^n:*

(i) Is $F(x,y) = F(x) + F(y)$ a joint cumulative distribution f^n?

(ii) Is $F(x,y) = F(x)F(y)$ a joint cumulative distribution f^n?

Sol. (*i*) Is not true because

$$\begin{aligned} F(\infty,\infty) &= F(\infty) + F(\infty) \\ &= 1 + 1 = 2 \neq 1 \end{aligned}$$

(*ii*) It is true because

(*a*) $F(\infty,\infty) = F(\infty)F(\infty) = 1.1 = 1$

(*b*) $F(-\infty,-y) = F(-\infty)F(-y) = 0$

$F(x,-\infty) = F(x)F(-\infty) = 0$

(*c*) $\dfrac{\partial^2 F}{\partial x\,\partial y} = \dfrac{\partial F(x)}{\partial x}\dfrac{\partial F(y)}{\partial y}$

$= f(x)\, g(y)$

(*d*) $F(x,y)$ is monotonic non-decreasing as $F(x), F(y)$ are so.

Ex. 11-10. *If x and y are two random variables having joint density function*

$$f(x,y) = \frac{1}{8}(6 - x - y),\ 0 < x < 2,\ 2 < y < 4.$$

Find (*i*) $p\{x \le 1, y < 3\}$

(*ii*) $p\{x + y < 3\}$

(*iii*) $p\{x < 1 / y < 3\}$.

Solution.

(*i*)
$$\begin{aligned} P\{x \le 1, y < 3\} &= \frac{1}{8}\int_{x=0}^{1}\int_{y=2}^{3}(6 - x - y)\,dy\,dx \\ &= \frac{1}{8}\int_0^1 dx \int_2^3 (6 - x - y)\,dy \\ &= \frac{1}{8}\int_0^1 dx\left\{6 - x - \frac{5}{2}\right\} \\ &= \frac{1}{8}\left\{6 - \frac{1}{2} - \frac{5}{2}\right\} = \frac{3}{8} \end{aligned}$$

(*ii*)
$$\begin{aligned} p(x + y < 3) &= \frac{1}{8}\int_{x=0}^{1}\int_{y=2}^{3-x}(6 - x - y)\,dx\,dy \\ &= \frac{1}{8}\int_0^1 dx\left\{6(1-x) - x(1-x) - \frac{1}{2}(5 + x^2 - 6x)\right\} \end{aligned}$$

$$= \frac{1}{8}\int_0^1 dx\left\{\frac{x^2}{2} - 4x + \frac{7}{2}\right\}$$

$$= \frac{1}{8}\left\{\frac{1}{6} - 2 + \frac{7}{2}\right\} = \frac{5}{24}$$

(*iii*) $$P\{x < 1/y < 3\} = \frac{P(x < 1, y < 3)}{P\{y < 3\}}$$

$$p(y < 3) = \int_{x=0}^{2}\int_{2}^{3}\frac{1}{8}(6 - x - y)\,dx\,dy$$

$$= \frac{1}{8}\int_0^2 dx\left\{6 - x - \frac{5}{2}\right\}$$

$$= \frac{1}{8}(12 - 2 - 5)$$

$$= \frac{5}{8}$$

$\therefore$ $$P\{x < 1/y < 3\} = \frac{3/8}{5/8} = 3/5.$$

Ex. 11-11. *Let x and y have the joint density function*

$$f(x, y) = \frac{1}{2}, 0 \le y \le x \le 2$$

Find the marginal and conditional probability density functions. Are x and y independent ?

Solution.

$$\int_{x=0}^{2}\int_{y=0}^{x} f(x, y)\,dx\,dy = \frac{1}{2}\int_{x=0}^{2} dx \int_{y=0}^{x} dy$$

$$= \frac{1}{2}\int_0^2 x\,dx$$

$$= 1$$

(*i*) $$f_x(x) = \int_{y=0}^{x} f(x, y)\,dy$$

$$= \frac{1}{2}\int_0^x dy = \frac{1}{2}x$$

$$f_y(y) = \int_y^2 f(x, y)\,dx$$

(*ii*) $f(x/$

$f(y/$

(*iii*) Now, $f_x(x)f_y($

$\therefore$ x and y are not independe

Ex. 11-12. *The joint density*

$f_{X,Y}(x,$

Determine marginal distribu

Solution.

$$\int_{x=0}^{2}\int_{y-0}^{1} f_{X,Y}(x, y)dx$$

(*i*) $f_X($

$f_Y($

(*ii*) Now $f_X(x)f_Y($

$\therefore$ X and Y are stochastically

$x+\frac{7}{2}\}$

$=\frac{5}{24}$

$y)\,dx\,dy$

$\frac{5}{2}\}$

on

≤ 2

density functions. Are x and y

$$= \frac{1}{2}\int_y^2 dx = \frac{1}{2}(2-y)$$

(ii)
$$f(x/y) = \frac{f(x,y)}{f_y(y)} = \frac{1}{2-y}$$

$$f(y/x) = \frac{f(x,y)}{f_x(x)} = \frac{1}{x}$$

(iii) Now,
$$f_x(x)f_y(y) = \frac{1}{2}x.\frac{1}{2}(2-y)$$

$$= \frac{1}{4}x(2-y) \neq f(x,y)$$

$\therefore$ x and y are not independent.

Ex. 11-12. *The joint density function of a bivariate distribution is given as below :*

$$f_{X,Y}(x,y) = \frac{1}{3}(x+y), 0 \leq x \leq 2, 0 \leq y \leq 1$$
$$= 0 \quad \textit{elsewhere}$$

Determine marginal distributions and show that X and Y are stochastically dependent.

Solution.

$$\int_{x=0}^{2}\int_{y=0}^{1} f_{X,Y}(x,y)\,dx\,dy = \frac{1}{3}\int_0^2 dx\int_0^1 (x+y)dy$$

$$= \frac{1}{3}\int_0^2 dx\left\{x+\frac{1}{2}\right\}$$

$$= \frac{1}{3}\left\{\frac{x^2}{2}+\frac{1}{2}x\right\} = 1$$

(i)
$$f_X(x) = \frac{1}{3}\int_0^1 (x+y)dy$$

$$= \frac{1}{3}\left(x+\frac{1}{2}\right)$$

$$f_Y(y) = \frac{1}{3}\int_0^2 (x+y)dx = \frac{2}{3}(1+y)$$

(ii) Now
$$f_X(x)f_Y(y) = \frac{2}{9}\left(x+\frac{1}{2}\right)(1+y) \neq f_{X,Y}(x,y)$$

$\therefore$ X and Y are stochastically dependent.

Ex. 11-13. *The joint density function of a bivariate distribution is given by*

$$f(x,y) = 4xy\, e^{-(x^2+y^2)},\ x \geq 0,\ y \geq 0.$$

Find the marginal and conditional probability density functions. Are x and y independent ?

Solution.

$$\int_0^\infty \int_0^\infty f(x,y)\,dx\,dy = 4\int_{x=0}^\infty \int_0^\infty xy\,e^{-(x^2+y^2)}\,dx\,dy$$

$$= 4\int_0^\infty dx.xe^{-x^2}\int_0^\infty ye^{-y^2}\,dy$$

$$= \left|-e^{-x^2}\right|_0^\infty \left|-e^{-y^2}\right|_0^\infty = 1$$

(i)
$$f_x(x) = \int_{y=0}^\infty 4xye^{-(x^2+y^2)}dy$$

$$= 4xe^{-x^2}\int_{y=0}^\infty ye^{-y^2}dy$$

$$= 2xe^{-x^2}\left|-e^{-y^2}\right|_0^\infty = 2xe^{-x^2}$$

$$f_y(y) = \int_{x=0}^\infty 4xye^{-(x^2+y^2)}dx$$

$$= 2ye^{-y^2}$$

(ii)
$$f_{x/y}(x/y) = \frac{f(x,y)}{f_y(y)} = \frac{4xye^{-(x^2+y^2)}}{2ye^{-y^2}}$$

$$= 2xe^{-x^2}$$

$$f_{y/x}(y/x) = \frac{4xye^{-(x^2+y^2)}}{2xe^{-x^2}} = 2ye^{-y^2}$$

(iii)
$$f(x,y) = 4xye^{-(x^2+y^2)}$$

$$= f_x(x)\,f_y(y)$$

$\therefore$ x and y are independent.

Ex. 11-14. *The joint density function of a bivariate distribution is given by*

$$f(x,y) = c\sin\frac{\pi}{2}(x+y), 0 < x < 1, 0 < y < 1$$
$$= 0 \text{ elsewhere.}$$

Find c and marginal; condition

Sol. c is given by

$\therefore$

(i) $f_x($

$f_y($

stribution is given by

$y \geq 0.$

ensity functions. Are x and y

dx dy

$^2 dy$

$= 1$

$2xe^{-x^2}$

$+y^2)$

2

e^{-y^2}

stribution is given by

$0 < y < 1$

Find c and marginal; conditional probability density functions. Are x and y independent ?

Sol. c is given by

$$1 = \int_{x=0}^{1} \int_{0}^{1} f(x,y)\,dx\,dy$$

$$= c \int_{x=0}^{1} dx \int_{0}^{1} \sin \frac{\pi}{2}(x+y)\,dy$$

$$= c\int_{0}^{1} dx \left\{ -\frac{2}{\pi} \cos \frac{\pi}{2}(x+y) \right\}_{0}^{1}$$

$$= \frac{2c}{\pi} \int_{0}^{1} \left\{ \cos \frac{\pi}{2} x - \cos \frac{\pi}{2}(1+x) \right\} dx$$

$$= \frac{4c}{\pi^2} \left| \sin \frac{\pi}{2} x - \sin \frac{\pi}{2}(1+x) \right|_{0}^{1}$$

$$= \frac{8}{\pi^2} c.$$

$$\therefore \qquad c = \frac{\pi^2}{8}$$

(*i*)
$$f_x(x) = \int_{0}^{1} f(x,y)\,dy$$

$$= c \int_{0}^{1} \sin \frac{\pi}{2}(x+y)\,dy$$

$$= \frac{\pi}{4} \left| -\cos \left(\frac{\pi}{2} \right)(x+y) \right|_{0}^{1}$$

$$= \frac{\pi}{4} \left\{ \cos \frac{\pi}{2} x - \cos \frac{\pi}{2}(1+x) \right\}$$

$$= \frac{\pi}{4} \left\{ \cos \frac{\pi}{2} x + \sin \frac{\pi}{2} x \right\}$$

$$f_y(y) = \int_{0}^{1} f(x,y)\,dx = \frac{\pi^2}{8} \int_{0}^{1} \sin \frac{\pi}{2}(x+y)\,dx$$

$$= \frac{\pi}{4} \left\{ \cos \frac{\pi}{2} y - \cos \frac{\pi}{2}(1+y) \right\}$$

$$= \frac{\pi}{4}\left\{\cos\frac{\pi}{2}y + \sin\frac{\pi}{2}y\right\}$$

(ii) $$f_{y/x}(y/x) = \frac{f(x,y)}{f_x(x)} = \frac{c.\sin\frac{\pi}{2}(x+y)}{\frac{\pi}{4}\left\{\cos\frac{\pi}{2}x + \sin\frac{\pi}{2}x\right\}}$$

$$= \frac{\pi}{2}\,\frac{\sin\frac{\pi}{2}(x+y)}{\cos\frac{\pi}{2}x + \sin\frac{\pi}{2}x}$$

$$f_{x/y}(x/y) = \frac{\pi}{2}\,\frac{\sin\frac{\pi}{2}(x+y)}{\cos\frac{\pi}{2}y + \sin\frac{\pi}{2}y}$$

(iii) Now $f(x,y) \neq f_x(x)\,f_y(y)$

$\therefore$ x and y are not independent.

Ex. 11-15. *Show that the conditions for the function*

$$f(x,y) = k\exp\{ax^2 + 2hxy + by^2\}, -\infty < x, y < \infty$$

be a density function are

(i) $a \le 0$ (ii) $b \le 0$ (iii) $ab - h^2 \ge 0$.

Assuming these conditions to be satisfied, find k

Sol. Let $f(x,y)$ be a density function. Then

$$\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

i.e., $$k\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty}\exp\left\{ax^2 + 2hxy + by^2\right\}dx\,dy = 1 \qquad \ldots(1)$$

Now $ax^2 + 2hxy + by^2$

$$= a\left\{\left(x + \frac{hy}{a}\right)^2 + \left(\frac{ab-h^2}{a^2}\right)y^2\right\}, \text{ if } a \neq 0 \text{ and}$$

$$= b\left\{\left(y + \frac{hx}{b}\right)^2 + \left(\frac{ab-h^2}{b^2}\right)x^2\right\}, \text{ if } b \neq 0$$

$\therefore$ The integral in (1) converges if

$a \le 0,\ b \le 0,\ ab - h^2 \ge 0$

Assume $a < 0, b < 0, ab - h^2 > 0$

Let $a = -\lambda, b = -\mu, h = \eta$.

Then $\lambda > 0, \mu > 0$ and $ab - h^2$

$\therefore$ $ax^2 + 2hxy + by^2$

$\therefore$ (1) $\Rightarrow$

$$k\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty}\exp\left[-\frac{(x\lambda - \eta y)^2}{\lambda} -\right.$$

i.e., $$1 = k\int_{y=-\infty}^{\infty}\left[\exp\left\{-\frac{\lambda\mu - \eta^2}{\lambda}y\right.\right.$$

Now $$\int_{x=-\infty}^{\infty}\exp.\left\{-\frac{(x\lambda - \eta y)^2}{\lambda}\right\}$$

$\therefore$ 1

$\therefore$ k

Assume $a < 0, b < 0, ab - h^2 > 0$

Let $a = -\lambda, b = -\mu, h = \eta.$

Then $\lambda > 0, \mu > 0$ and $ab - h^2 = \lambda\mu - \eta^2 > 0$

$$\therefore \quad ax^2 + 2hxy + by^2 = -\lambda\left\{\left(x - \frac{\eta}{\lambda}y\right)^2 + \frac{\lambda\mu - \eta^2}{\lambda_2}y^2\right\}$$

$$= -\frac{(x\lambda - \eta y)^2}{\lambda} - \frac{\lambda\mu - \eta^2}{\lambda}y^2.$$

$\therefore (1) \Rightarrow$

$$k\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty}\exp\left[-\frac{(x\lambda - \eta y)^2}{\lambda} - \frac{\lambda\mu - \eta^2}{\lambda}y^2\right]dx\,dy = 1$$

$$i.e., \; 1 = k\int_{y=-\infty}^{\infty}\left[\exp\left\{-\frac{\lambda\mu - \eta^2}{\lambda}y^2\right\}\int_{x=-\infty}^{\infty}\left\{-\frac{(x\lambda - \eta y)^2}{\lambda}\right\}dx\right]dy$$

Now $$\int_{x=-\infty}^{\infty}\exp.\left\{-\frac{(x\lambda - \eta y)^2}{\lambda}\right\}dx = \frac{1}{\lambda}\int_{-\infty}^{\infty}\exp\left(-\frac{u^2}{\lambda}\right)du \text{ where } u = x\lambda - \eta y$$

$$= \frac{1}{\sqrt{\lambda}}\int_0^{\infty} t^{\frac{1}{2}-1}e^{-t}\,dt \qquad \text{where } u^2 = \lambda t$$

$$= \frac{1}{\sqrt{\lambda}}\,\Gamma\left(\frac{1}{2}\right)$$

$$= \frac{\sqrt{\pi}}{\sqrt{\lambda}}$$

$$\therefore \quad 1 = k.\frac{\sqrt{\pi}}{\sqrt{\lambda}}\int_{y=-\infty}^{\infty}\exp\left\{-\frac{\lambda\mu - \eta^2}{\lambda}y^2\right\}dy$$

$$= k\frac{\sqrt{\pi}}{\sqrt{\lambda}}.\sqrt{\pi}\frac{\sqrt{\lambda}}{\sqrt{\lambda\mu - \eta^2}}$$

$$\therefore \quad k = \frac{1}{\pi}\sqrt{\lambda\mu - \eta^2}$$

$$= \frac{1}{\pi}\sqrt{ab - h^2}$$

11.3. Joint Moment Generating Function and Moments

Def. 1. For random variates x and y, the joint moments about point (a, b) is defined to be

$$E\{(x-a)^r (y-b)^s\}$$

where r, s are zero or any positive integers. It is denoted by $\mu_{rs}'(a, b)$.

If $a = \bar{x}, b = \bar{y}$ joint moment is denoted by μ_{rs}. Thus

$$\mu_{rs} = E\{(x-\bar{x})^r (y-\bar{y})^s\}$$

Def. 2. For random variates x and y the joint moment generating function is defined by

$$M_{a,b}(t_1,t_2) = E\{e^{t_1(x-a)+t_2(y-b)}\}$$

where $M_{a,b}(t_1,t_2)$ denotes *m.g.f.* about point (a, b).

Here it is assumed that t_1, t_2 are real and expectation exists for all values of t_1, t_2 such that $-h < t_1, t_2 < h$ for some $h > 0$.

The relation between joint moments and joint *m.g.f.* is

$$\mu_{rs}' = \left[\frac{\partial^{r+s}}{\partial t_1^r\, \partial t_2^s}\{M_{a,b}(t_1,t_2)\}\right]_{t_1=0=t_2}$$

and μ_{rs}' is the co-eff. of $\dfrac{t_1^r}{r!}\dfrac{t_2^s}{s!}$ in the expansion of $M_{a,b}(t_1,t_2)$.

In general, if $x_1, x_2, \ldots\ldots x_n$ are n variates, joint moment about point $(a_1, a_2 \ldots\ldots a_n)$ is defined by

$$\mu_{r_1,r_2\ldots r_n}'(a_1, a_2, \ldots a_n) = E\{(x_1-a_1)^{r_1}(x_2-a_2)^{r_2}\ldots(x_n-a_n)^{r_n}\}$$

and joint moment generating f^n about point $(a_1, a_2, \ldots\ldots a_n)$ is defined by

$$M_{a_1,a_2,\ldots\ldots a_n}(t_1,t_2,\ldots\ldots t_n) = E\left\{e^{\sum_{i=1}^{n} t_i(x_i-a_i)}\right\}$$

In order to obtain $\mu_{r_1,r_2,\ldots r_n}$ *m.g.f.* is differentiated r_1 times *w.r.t.* t_1, r_2 times *w.r.t.* t_2 and so on and then limit is taken as all t's approach 0.

or co-eff. $\dfrac{t_1^{r_1}}{r_1!}\cdot\dfrac{t_2^{r_2}}{r_2!}\cdots\dfrac{t_n^{r_n}}{r_n!}$ is taken in the expansion of m.g.f.

Remark : (1) Marginal moment generating functions can be obtained from joint m.g.f. as below :

$$M_a(t_1) = M_{a,b}(t_1, 0) = \lim_{t_2 \to 0} M_{a,b}(t_1, t_2)$$

$$M_b(t_2) = M_{a,b}(0, t_2) = \lim_{t_1 \to 0} M_{a,b}(t_1, t_2)$$

(2) The independence of two
the help of *m.g.fs*. The result is :
Two jointly distributed randc

$M_{x+y}(t_1,$

where $M_{x+y}(t_1, t_2)$ denotes joint
The proof of this result is be

11.4. Conditional Expectation

Consider a bivariate distribu
of x and y.
Conditional expectation of '

and

It is denoted by

Similarly condition expectat
Theorem : 11.4.1. $E[g($
Proof :
Let h

Then $E[E\{g(y)/$

(2) The independence of two jointly distributed variates x and y can also be tested with the help of $m.g.fs$. The result is :

Two jointly distributed random variates x and y are independent if and only if

$$M_{x+y}(t_1, t_2) = M_x(t_1).M_y(t_2)$$

where $M_{x+y}(t_1, t_2)$ denotes joint $m.g.f$ about (0, 0) etc.

The proof of this result is beyond the scope of this book.

11.4. Conditional Expectation

Consider a bivariate distribution with variates x and y. Let $g(x, y)$ be a continuous f^n of x and y.

Conditional expectation of 'g' given x is defined to be

$$\sum_j g(x, y_j) p(y_j / x) \qquad \text{for discrete distribution}$$

and

$$\int_{-\infty}^{\infty} g(x, y) f(y / x)\, dy. \qquad \text{for continuous dist.}$$

It is denoted by

$$E\{g(x, y) / x\}$$

Similarly condition expectations given y can be defined.

Theorem : 11.4.1. $\quad E[g(y)\} = E[E\{g(y)/x\}] \qquad \ldots(1)$

Proof :

Let
$$h(x) = E\{g(y)/x)$$
$$= \sum_j g(y_j) p(y_j / x) \qquad \ldots(i)$$

Then
$$E[E\{g(y)/x\}] = E[(h(x)]$$
$$= \sum_i h(x_i) p_x(x_i)$$
$$= \sum_i \left\{ \sum_j g(y_j) p(y_j / x_i) \right\} p_x(x_i) \qquad \text{(using } (i))$$
$$= \sum_i \sum_j g(y_j) p(y_j / x_i) p_x(x_i)$$
$$= \sum_i \sum_j g(y_j) p_{ij}$$
$$= E\{g(y)\} \qquad \ldots(2)$$

Remark : (*i*) Similarly as above it can be shown that

$$E[E\{g(x)/y\}] = E\{g(x)\} \qquad \ldots(3)$$

(*ii*) If $g(y) = y$ then (1) $\Rightarrow$

$$E[E(y/x)] = E(y) \qquad \ldots(4)$$

and if $g(x) = x$, (3) $\Rightarrow$

$$E[E(x/y)] = E(x) \qquad \ldots(5)$$

Result 11.4.2. If g_1, g_2 are f^n's of one variable then

$$E\{g_1(y)+g_2(y)/x\} = E\{g_1(y)/x\}+E\{g_2(y)/x)$$

$$E\{g_1(y)g_2(x)/x) = g_2(x)E\{g_1(y)/x\}$$

Similar results hold when y is given.

Ex. 11.4.3. Result : *Conditional mean coincides with unconditional mean if and only if the variates are independent.*

Sol. Let x and y be the variates with joint density $f^n f(x,y)$.

Now
$$E(y/x) = \int y\, f_{y/x}(y/x)dy$$

$$= \int y \frac{f(x,y)}{f_x(x)} dy$$

Also
$$E(y) = \int y\, f_y(y)dy$$

Where the integrals are extended over the respective range

$\therefore$ $E(y/x)$ and $E(y)$ coincides iff.

$$\frac{f(x,y)}{f_x(x)} = f_y(y)$$

i.e.,
$$f(x,y) = f_x(x)f_y(y)$$

$\Rightarrow$ x and y are independent.

11.5. Conditional Variance

Conditional Variance of y given x is denoted by

$$\text{var}[y/x]$$

and is defined by

$$\text{var}\,\{y/x\} = E\left[\{y-E(y/x)\}^2/x\right]$$

$$= E\left[\left\{y^2-2yE(y/x)+\{E(y/x)\}^2\right\}/\{x\}\right]$$

$$= E(y^2/x)-\{E(y/x)\}^2 \qquad \ldots(1)$$

Similarly conditional variance of x given y is defined.

Theorem 11.5.1. $\text{var}\,(y) = E[\text{var}(y/x)]+\text{var}\,[E(y/x)]$

Proof : By (1) we have

$$\text{var}(y/x) = E(y^2/x)-\{E(y/x)\}^2$$

$\Rightarrow$ $E\left[\text{var}(y/x)\right]$

Also $\text{var}\left[E(y/x)\right]$

Adding (2) and (3)

$E\left[\text{var}(y/x)\right]+\text{var}\left[E(y/x)\right]$

Ex. 11-16. For the bivariate dis

$f(x, y) =$

find the conditional mean and variai

Sol. We have

1

$\Rightarrow$ y_0

$\therefore$ $f(x,y)$

Marginal density f^n of y is gi

$f_y(y)$

$\therefore$ Conditional density f^n of

$f_{x/y}(x/y)$

...(3)

...(4)

...(5)

$(y)/x)$

unconditional mean if and only

(x, y).

ange

$x]$

$+\{E(y/x)\}^2\}/\{x)\}]$

$)\}^2$ ...(1)

$)]$

$)\}^2$

$$\Rightarrow \qquad E[\text{var}(y/x)] = E[E(y^2/x)] - E\{E(y/x)\}^2$$

$$= E(y^2) - E\{E(y/x)\}^2 \qquad \ldots(2)$$

Also

$$\text{var}[E(y/x)] = E\{E(y/x)\}^2 - [E\{E(y/x)\}]^2$$

$$= E\{E(y/x)\}^2 - \{E(y)\}^2 \qquad \ldots(3)$$

using (4) of 11.4.1.

Adding (2) and (3)

$$E[\text{var}(y/x)] + \text{var}[E(y/x)] = E(y^2) - \{E(y)\}^2$$

$$= \text{var}(y).$$

Ex. 11-16. For the bivariate distribution

$$f(x, y) = y_0 x^2 y^3, \qquad 0 < x < y < 1,$$

find the conditional mean and variance of x for given y.

Sol. We have

$$1 = \int\int f(x, y)\,dx\,dy$$

$$= y_0 \int_{y=0}^{1} dy \int_0^y x^2 y^3\,dx$$

$$= \frac{y_0}{3} \int_{y=0}^{1} y^6\,dy = \frac{y_0}{21}$$

$$\Rightarrow \qquad y_0 = 21$$

$$\therefore \qquad f(x, y) = 21x^2y^3, 0 < x < y < 1.$$

Marginal density f^n of y is given by

$$f_y(y) = \int_0^y f(x, y)\,dx$$

$$= 21y^3 \int_0^y x^2\,dx$$

$$= 7y^6$$

$\therefore$ Conditional density f^n of x for given y is given by

$$f_{x/y}(x/y) = \frac{f(x, y)}{f_y(y)}$$

$$= \frac{3x^2}{y^3}$$

$\therefore$ Conditional mean of x for given y is given by

$$E(x/y) = \int_0^y x.\frac{3x^2}{y^3}dx$$

$$= \frac{3}{4}y$$

Also
$$E(x^2/y) = \int_0^y x^2.\frac{3x^2}{y^3}dx$$

$$= \frac{3}{5}y^2$$

$\therefore$
$$\text{var}(x/y) = E(x^2/y) - \{E(x/y)\}^2$$

$$= \frac{3}{5}y^2 - \frac{9}{16}y^2$$

$$= \frac{3}{80}y^2.$$

Ex. 11-17. *Three onbiased coins are tossed. X denotes the number of heads on the first two and Y denote the number of heads on the last two. Find :*

(i) the joint distribution of X and Y

(ii) $E\{Y/X=1\}$

(iii) $\rho_{X,Y}$

(iv) Give a joint distribution that is not the joint distribution given in part (i) yet has the same marginal distributions as the joint distribution given in (i).

Sol. Different possibilities are :

HHH, HHT, HTH, HTT
THH, THT, TTH, TTT

Possible values of X, Y each are 0, 1, 2.

$\therefore$ Joint dist. is

$Y \downarrow$ \ $X \rightarrow$	0	1	2	$f_Y(y)$
0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{2}{8}$
1	$1/8$	$2/8$	$1/8$	$4/8$
2	0	$\frac{1}{8}$	$1/8$	$2/8$
$f_X(x)$	$2/8$	$4/8$	$2/8$	

$E\{Y/X=1\}$

$E(XY)$

$E(X)$

$E(X^2)$

var $(X$

Similarly var $(Y$

cov $(X,Y$

$\therefore$ ρ_x

(*iv*) Reqd. joint distribution is

$f_{X,Y}(x,y$

$\therefore$ Joint dist. is :

$Y \downarrow$ \ $X \rightarrow$	0
0	$\frac{2}{8}\cdot\frac{2}{8} = \frac{4}{64}$
1	$\frac{2}{8}\cdot\frac{4}{8} = \frac{8}{64}$
2	$\frac{2}{8}\cdot\frac{2}{8} = \frac{4}{64}$
	1/4

$/ y)\}^2$

tes the number of heads on the first
'ind :

stribution given in part (i) yet has
given in (i).

2	$f_Y(y)$
0	$\frac{2}{8}$
$\frac{1}{8}$	$\frac{4}{8}$
$\frac{1}{8}$	$\frac{2}{8}$
$\frac{2}{8}$	

$$E\{Y / X = 1\} = \left(0 \cdot \frac{1}{8} + 1 \cdot \frac{2}{8} + 2 \cdot \frac{1}{8}\right) \Big/ 4/8$$

$$= 1.$$

$$E(XY) = \frac{1}{8} \cdot 0 \cdot 0 + \frac{1}{8} 1 \cdot 0 + 0 \cdot 2 \cdot 0 + \frac{1}{8} \cdot 0 \cdot 1 + \frac{2}{8} 1 \cdot 1 + \frac{1}{8} \cdot 2 \cdot 1$$

$$+ 0 \cdot 0 \cdot 2 + \frac{1}{8} \cdot 1 \cdot 2 + \frac{1}{8} \cdot 2 \cdot 2$$

$$= \frac{2}{8} + \frac{2}{8} + \frac{2}{8} + \frac{4}{8} = \frac{5}{4}$$

$$E(X) = 0 . \frac{2}{8} + 1 . \frac{4}{8} + 2 . \frac{2}{8} = 1$$

$$E(X^2) = 0^2 . \frac{2}{8} + 1^2 . \frac{4}{8} + 2^2 . \frac{2}{8} = \frac{12}{8} = \frac{3}{2}$$

$$\text{var}(X) = \frac{3}{2} - 1 = \frac{1}{2}.$$

Similarly $\text{var}(Y) = \frac{1}{2}$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$= \frac{5}{4} - 1 = \frac{1}{4}$$

$$\therefore \quad \rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) . \text{var}(y)}} = \frac{1/4}{1/2} = \frac{1}{2}$$

(*iv*) Reqd. joint distribution is given by

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

$\therefore$ Joint dist. is :

$X \rightarrow$ / $Y \downarrow$	0	1	2	
0	$\frac{2}{8} \cdot \frac{2}{8} = \frac{4}{64}$	$\frac{4}{8} \cdot \frac{2}{8} = \frac{8}{64}$	$\frac{2}{8} \cdot \frac{2}{8} = \frac{4}{64}$	$\frac{1}{4}$
1	$\frac{2}{8} \cdot \frac{4}{8} = \frac{8}{64}$	$\frac{4}{8} \cdot \frac{4}{8} = \frac{16}{64}$	$\frac{2}{8} \cdot \frac{4}{8} = \frac{8}{64}$	$\frac{1}{2}$
2	$\frac{2}{8} \cdot \frac{2}{8} = \frac{4}{64}$	$\frac{4}{8} \cdot \frac{2}{8} = \frac{8}{64}$	$\frac{2}{8} \cdot \frac{2}{8} = \frac{4}{64}$	$\frac{1}{4}$
	1/4	1/2	1/4	1

Ex. 11-18. *Let a random variable X has a density function* $f_X(\cdot)$ *and cumulative distribution function* $F_X(\cdot)$, *mean* μ *and s.d.* σ. *Let*

$$Y = \alpha + \beta X, \qquad -\infty < \alpha < \infty, \quad \beta > 0$$

Then :

(*i*) *find* α, β *suth that Y has zero mean and variance 1.*

(*ii*) *what is the correlation coefficient between X and Y ?*

(*iii*) *find cumulative distribution function of Y in terms of* α, β *and* $F_X(\cdot)$

(*iv*) *if X is symmetrically distributed about* μ, *is Y necessarily symmetrically distributed about its mean ?*

Sol. (*i*) $Y = \alpha + \beta X \Rightarrow E(Y) = \alpha + \beta\, E(X)$

$\therefore\ 0 = \alpha + \beta\mu \Rightarrow \alpha = -\beta\mu.$ $\qquad (\because E(y) = 0)$

$$\begin{aligned} 1 &= E(Y-\overline{Y})^2 \\ &= E\left\{\beta^2 (X-\overline{X})^2\right\} \\ &= \beta^2 . \sigma^2 \end{aligned}$$

$$\therefore \qquad \beta = \frac{1}{\sigma}$$

$$\therefore \qquad \alpha = -\frac{\mu}{\sigma}.$$

(*ii*)

$$\begin{aligned} \text{cov}(X,Y) &= E\left\{(X-\overline{X})(Y-\overline{Y})\right\} \\ &= \beta\, E(X-\overline{X})^2 = \beta\sigma^2 \\ \text{var}(Y) &= E(Y-\overline{Y})^2 \\ &= \beta^2 E(X-\overline{X})^2 \\ &= \beta^2\sigma^2 \end{aligned}$$

$$\therefore \qquad \rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X . \sigma_Y} = \frac{\beta\sigma^2}{\sigma\sqrt{\beta^2\sigma^\sigma}}$$

$$= \frac{\beta}{|\beta|} = \begin{cases} 1 & \text{if } \beta > 0 \\ -1 & \text{if } \beta < 0 \end{cases}$$

(*iii*) *c.d.f.* of *Y* is given by

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(\alpha + \beta X \le y) \\ &= P\left(X \le \frac{y-\alpha}{\beta}\right) \end{aligned}$$

(*iv*) We have : if *Z* is symm
have the same distribution.

$\therefore\ X - \mu$ and $-(X-\mu)$ hav
Now $Y - \overline{Y} = \beta(X - \overline{X}) = \beta($
$\therefore\ Y - \overline{Y}$ and $-(Y-\overline{Y})$ als
$\therefore\ Y$ is symmetrically distr

Ex. 11-19. *Three fair coins*
two coins and Y denotes the num

Find: (*i*) *The joint distrib*
(*ii*) *Conditional dis*
(*iii*) cov (X,Y).

Sol. Different possibilities a

$\therefore$ Possible values X, Y are
$X = 2, Y = 0$
$X = 1, Y = 0\,;$

$\therefore$ Joint density function is

$X \rightarrow$ / $Y \downarrow$	
0	
1	
2	
$f_X(x)$	

Conditional dist. of Y give

$P(Y = 0 | X$

$P(Y = 1 | X$

$P(Y = 2 | X$

'ty function $f_Y(\cdot)$ and cumulative

$\beta > 0$

'ance 1.

and Y ?

erms of α, β and $F_X(\cdot)$

, is Y necessarily symmetrically

$(\because E(y) = 0)$

)}

σ^2

σ^2

$\beta^2\sigma^\sigma$

> 0

< 0

$$= F_X\left(\frac{y-\alpha}{\beta}\right)$$

(*iv*) We have : if Z is symmetrically distributed about constant $c, Z - c$ and $-(Z - c)$ have the same distribution.

$\therefore$ $X - \mu$ and $-(X - \mu)$ have the same distribution.

Now $Y - \overline{Y} = \beta(X - \overline{X}) = \beta(X - \mu)$.

$\therefore$ $Y - \overline{Y}$ and $-(Y - \overline{Y})$ also have the same distribution.

$\therefore$ Y is symmetrically distributed about its mean $\overline{Y}$.

Ex. 11-19. *Three fair coins are tossed. Let X denote the number of heads on the first two coins and Y denotes the number of tails on the last two coins.*

Find: (i) The joint distribution of X and Y.

(ii) Conditional distribution of Y given X=1.

(iii) cov (X,Y).

Sol. Different possibilities are

HHH, HHT, HTH, HTT
THH, THT, TTH, TTT

$\therefore$ Possible values X, Y are :

$$X = 2, Y = 0;\ X = 2, Y = 1;\ X = 1, Y = 1;\ X = 1, Y = 2$$

$$X = 1, Y = 0;\ X = 1, Y = 1;\ X = 0, Y = 1;\ X = 0, Y = 2$$

$\therefore$ Joint density function is :

$Y \downarrow$ \ $X \rightarrow$	0	1	2	$f_Y(y)$
0	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$
1	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{4}{8}$
2	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{2}{8}$
$f_X(x)$	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	

Conditional dist. of Y given $X = 1$ is as below :

$$P(Y = 0 \mid X = 1) = \frac{f_{X,Y}(0,1)}{f_X(1)} = \frac{1/8}{4/8} = \frac{1}{4}$$

$$P(Y = 1 \mid X = 1) = \frac{f_{X,Y}(1,1)}{f_X(1)} = \frac{2/8}{4/8} = \frac{2}{4}$$

$$P(Y = 2 \mid X = 1) = \frac{1/8}{4/8} = \frac{1}{4}$$

(ii)
$$E(XY) = 0\cdot0\cdot0+0\cdot1\cdot\frac{1}{8}+0\cdot2\cdot\frac{1}{8}+1\cdot0\cdot\frac{1}{8}+1\cdot1\cdot\frac{2}{8}$$
$$+1\cdot2\cdot\frac{1}{8}+2\cdot0\cdot\frac{1}{8}+2\cdot1\cdot\frac{1}{8}+2\cdot2\cdot0$$
$$= \frac{2}{8}+\frac{2}{8}+\frac{2}{8}=\frac{6}{8}$$
$$E(X) = 0\cdot\frac{2}{8}+1\cdot\frac{4}{8}+2\cdot\frac{2}{8}=\frac{8}{8}=1$$
$$E(Y) = 0\cdot\frac{2}{8}+1\cdot\frac{4}{8}+2\cdot\frac{2}{8}=1$$
$$\therefore \quad \text{cov}(X,Y) = E(XY)-E(X)\,E(Y)$$
$$= \frac{6}{8}-1\cdot1=-\frac{1}{4}.$$

Ex. 11-20. *Consider the experiment of tossing of two tetrahedrons. Let X be the number on the first and Y the larger of two numbers. Obtain joint discrete density function of X and Y. Also find:* $E(XY), E(X+Y), E(X), E(Y), E(X^2), E(Y^2)$, $\text{var}(X), \text{var}(Y), \text{cov}(X,Y)$ *and* $\gamma_{X,Y}$, $E(Y|X=1)$, $E(X|Y=2)$.

Sol. For joint discrete density function see Ex : 11-5

$$E(X) = 1\cdot\frac{4}{16}+2\cdot\frac{4}{16}+3\cdot\frac{4}{16}+4\cdot\frac{4}{16}=\frac{5}{2}$$
$$E(Y) = 1\cdot\frac{1}{16}+2\cdot\frac{3}{16}+3\cdot\frac{5}{16}+4\cdot\frac{7}{16}=\frac{50}{16}$$
$$E(X^2) = 1^2\cdot\frac{4}{16}+2^2\cdot\frac{4}{16}+3^2\cdot\frac{4}{16}+4^2\cdot\frac{4}{16}=\frac{4}{16}\{1+4+9+16\}=\frac{30}{4}$$
$$E(Y^2) = 1^2\cdot\frac{1}{16}+2^2\cdot\frac{3}{16}+3^2\cdot\frac{5}{16}+4^2\cdot\frac{7}{16}$$
$$= \frac{1}{16}\{1+12+45+112\}=\frac{170}{16}$$
$$E(XY) = 1\cdot1\cdot\frac{1}{16}+1\cdot2\cdot\frac{1}{16}+1\cdot3\cdot\frac{1}{16}+1\cdot4\cdot\frac{1}{16}+2\cdot2\cdot\frac{2}{16}+2\cdot3\cdot\frac{1}{16}+2\cdot4\cdot\frac{1}{16}$$
$$+3\cdot3\cdot\frac{3}{16}+3\cdot4\cdot\frac{1}{16}+4\cdot4\cdot\frac{4}{16}$$
$$= \frac{1}{16}\{1+2+3+4+8+6+8+27+12+64\}=\frac{135}{16}$$
$$E(X+Y) = \frac{1}{16}(1+1)+(1+2)\cdot\frac{1}{16}+(1+3)\cdot\frac{1}{16}+(1+4)\cdot\frac{1}{16}+(2+2)\cdot\frac{2}{16}+(2+3)\cdot\frac{1}{16}$$

$+(2+4)\frac{1}{16}+(3+$

$= \frac{1}{16}\{2+$

$\text{var}(X) = E(X^2)-$

$= \frac{30}{4}-\frac{25}{4}$

$\text{var}(Y) = E(Y^2)-$

$\text{cov}(X,Y) = E(XY)-$

$= \frac{135}{16}-\frac{5}{2}$

$\gamma_{xy} = \frac{\text{cov}(}{\sqrt{\text{var}(X}}$

$f_{Y/X}(y/1) = \frac{f_{X,Y}(x,}{f_X(1)}$

$f_{X/Y}(x/2) = \frac{f_{X,Y}(x,}{f_Y(2)}$

$\therefore \quad E(Y/X=1) = \frac{1}{4}\{1+2$

$E(X/Y=2) = \frac{1}{\frac{3}{16}}\{1\cdot$

Ex. 11-21. *For the bivariate di*

$f_{X,Y}(x,$

Find : $E(X), E(Y), E(X^2), E($

$\text{cov}(X,Y$

Sol. $f_X(x) = \int_0^1 (x+y$

$f_Y(y) = \int_0^1 (x+y$

$\frac{1}{8}+1\cdot 1\cdot\frac{2}{8}$

$2\cdot 0$

...trahedrons. Let X be the number ...iscrete density function of X and , $\text{var}(X), \text{var}(Y), \text{cov}(X,Y)$ *and*

$\frac{5}{2}$

$\frac{50}{16}$

$\frac{4}{6}=\frac{4}{16}\{1+4+9+16\}=\frac{30}{4}$

$\frac{7}{6}$

$\frac{2}{16}+2\cdot 3\cdot\frac{1}{16}+2\cdot 4\cdot\frac{1}{16}$

$+12+64\}=\frac{135}{16}$

$)\cdot\frac{1}{16}+(2+2)\cdot\frac{2}{16}+(2+3)\cdot\frac{1}{16}$

$$+(2+4)\frac{1}{16}+(3+3)\frac{3}{16}+(3+4)\frac{1}{16}+(4+4)\frac{4}{16}$$

$$= \frac{1}{16}\{2+3+4+5+8+5+6+18+7+32\}=\frac{90}{16}$$

$$\text{var}(X) = E(X^2)-\{E(X)\}^2$$

$$= \frac{30}{4}-\frac{25}{4}=\frac{5}{4}$$

$$\text{var}(Y) = E(Y^2)-\{E(Y)\}^2=\frac{170}{16}-\left(\frac{50}{16}\right)^2=\frac{55}{64}$$

$$\text{cov}(X,Y) = E(XY)-E(X)\,E(Y)$$

$$= \frac{135}{16}-\frac{5}{2}\cdot\frac{25}{8}=\frac{10}{16}$$ (See Chapter 13).

$$\gamma_{xy} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X).\text{var}(Y)}}=\frac{10/16}{\sqrt{\frac{5}{4}\cdot\frac{55}{64}}}=\frac{2}{\sqrt{11}}$$

$$f_{Y/X}(y/1) = \frac{f_{X,Y}(x,y)}{f_X(1)}$$

$$f_{X/Y}(x/2) = \frac{f_{X,Y}(x,y)}{f_Y(2)}$$

$$\therefore \quad E(Y/X=1) = \frac{1}{4}\{1+2+3+4\}=\frac{5}{2}$$

$$E(X/Y=2) = \frac{1}{\frac{3}{16}}\left\{1\cdot\frac{1}{16}+2\cdot\frac{2}{16}\right\}=5/3.$$

Ex. 11-21. *For the bivariate distribution given by*

$$f_{X,Y}(x,y)=(x+y)\,I_{(0,1)}(x)\,I_{(0,1)}(y)$$

Find : $E(X), E(Y), E(X^2), E(Y^2), E(XY), E(X+Y), \text{var}(X), \text{var}(Y)$

$\text{cov}(X,Y)$ *and* $\gamma_{X,Y}$, $E\{Y/X=x\}$.

Sol. $$f_X(x) = \int_0^1(x+y)\,dy=x+\frac{1}{2} \quad , \quad 0<x<1$$

$$f_Y(y) = \int_0^1(x+y)\,dx=y+\frac{1}{2} \quad , \quad 0<y<1$$

$$\therefore \qquad E(X) = \int_0^1 x\, f_X(x) = \int_0^1 x\left(x+\frac{1}{2}\right)dx = \frac{1}{3}+\frac{1}{4} = \frac{7}{12}$$

Similarly,
$$E(Y) = 7/12$$

$$E(X^2) = \int_0^1 x^2\left(x+\frac{1}{2}\right)dx = \frac{1}{4}+\frac{1}{6} = \frac{5}{12}$$

$$E(Y^2) = 5/12$$

$$\text{var}(X) = E(X^2)-\{E(X)\}^2 = \frac{5}{12}-\frac{49}{144} = \frac{11}{144}$$

also
$$\text{var}(Y) = \frac{11}{144}$$

$$E(XY) = \int_0^1\int_0^1 xy(x+y)\,dx\,dy$$

$$= \int_0^1 x\,dx\left\{\frac{x}{2}+\frac{1}{3}\right\} = \frac{1}{6}+\frac{1}{6} = \frac{1}{3}$$

$$E(X+Y) = \int_0^1\int_0^1 (x+y).(x+y)\,dx\,dy$$

$$= \int_0^1 dx \int_0^1 (x+y)^2\,dy$$

$$= \int_0^1 dx\left\{\frac{(x+y)^3}{3}\right\}_{y=0}^{1}$$

$$= \frac{1}{3}\int_0^1 dx\left\{(1+x)^3 - x^3\right\}$$

$$= \frac{1}{3}\left[\frac{(1+x)^4}{4}-\frac{x^4}{4}\right]_0^1 = \frac{1}{12}\{16-1-1\} = \frac{7}{6}$$

$$\text{cov}(X,Y) = E(XY)-E(X).E(Y)$$

$$= \frac{1}{3}-\frac{7}{12}\cdot\frac{7}{12} = -\frac{1}{144}$$

$\gamma_{X,Y}$

$f_{Y/X}(y/x$

$\therefore \qquad E\{Y/X=x\}$

Ex. 11-22. *Suppose that the ran*
(0,1) *i.e.,* $f_X(x) = I_{(0,1)}(x)$. *Also th*
with parameters n and x i.e.,

$P(Y=y|X=x$

Find : (i) $E(Y)$ *(ii)* Distributi

Sol. As conditional distributio

$E(Y|X=x$

$\therefore \qquad E(Y$

Joint density function of X and

$f_{X,Y}(x,y$

$\therefore$ Distribution of Y is

$f_Y(y$

$$\left(x+\frac{1}{2}\right)dx = \frac{1}{3}+\frac{1}{4} = \frac{7}{12}$$

$$x = \frac{1}{4}+\frac{1}{6} = \frac{5}{12}$$

$$)\}^2 = \frac{5}{12} - \frac{49}{144} = \frac{11}{144}$$

$dx\, dy$

$$\} = \frac{1}{6}+\frac{1}{6} = \frac{1}{3}$$

$+ y)\, dx\, dy$

dy

$$\Big\}_{y=0}^{1}$$

$$^3 - x^3\}$$

$$\frac{x^4}{4}\Bigg]_0^1 = \frac{1}{12}\{16-1\ \ \} = \frac{7}{6}$$

$.E(Y)$

$$-\frac{1}{144}$$

$$\gamma_{X,Y} = \frac{\operatorname{cov}(X.Y)}{\sqrt{\operatorname{var}(X).\operatorname{var}(Y)}} = \frac{-1/144}{\sqrt{\frac{11}{144}\cdot\frac{11}{144}}} = -\frac{1}{11}$$

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \frac{(x+y)}{x+\frac{1}{2}} I_{(0,1)}(y)$$

$$\therefore \qquad E\{Y/X=x\} = \int_0^1 \frac{y(x+y)}{x+\frac{1}{2}}\,dy$$

$$= \frac{1}{x+\frac{1}{2}}\left\{\frac{x}{2}+\frac{1}{3}\right\} \qquad 0<x<1.$$

Ex. 11-22. *Suppose that the random variable X is uniformly distributed over the interval* (0,1) *i.e.,* $f_X(x) = I_{(0,1)}(x)$. *Also the conditional distribution of Y given X = x is binomial with parameters n and x i.e.,*

$$P(Y=y|X=x) = {}^nc_y\, x^y(1-x)^{n-y},\ y = 0,1,\ldots\ldots\ldots n.$$

Find : (i) $E(Y)$ *(ii)* Distribution of Y.

Sol. As conditional distribution of Y is binomial with parameters $n;\ x$.

$$E(Y|X=x) = nx$$

$$\therefore \qquad E(Y) = E\{E(Y|X=x)\} = E(nx)$$

$$= nE(x)$$

$$= n\int_0^1 x\,dx = \frac{n}{2}$$

Joint density function of X and Y is

$$f_{X,Y}(x,y) = f_X(x)\,f_{Y/X}(y/x)$$

$$= {}^nc_y\, x^y(1-x)^{n-y} \qquad y = 0,1,2\ldots\ldots n.$$

$$0 \le x \le 1$$

$\therefore$ Distribution of Y is

$$f_Y(y) = \int_0^1 {}^nc_y\, x^y(1-x)^{n-y}\,dx$$

$$= {}^nc_y \int_0^1 x^{y+1-1}(1-x)^{n-y+1-1}\,dx$$

$$= {}^n c_y . \beta(y+1, n-y+1)$$

$$= {}^n c_y \frac{y+1 \quad n-y+1}{(y+1+n-y+1)}$$

$$= {}^n c_y \frac{\Gamma(y+1)\,\Gamma.(n-y+1)}{\Gamma(n+2)}$$

$$= \frac{n!}{y!(n-y)!} \frac{y! \ \ (n-y)!}{(n+1)!} = \frac{1}{n+1} \qquad n = 0, 1, 2, ...n.$$

Ex. 11-23. *For the distribution given by*

$$f_{X,Y}(x,y) = e^{-x-y}\, I_{(0,\infty)}(x)\, I_{(0,\infty)}(y)$$

Find : (*i*) $P(X > 1)$ (*ii*) $P(1 < X + Y < 2)$ (*iii*) $P(X < Y | X < 2y)$

(*iv*) *m. s.t* $P(X + Y < m) = \frac{1}{2}$ (*v*) $P(0 < X < 1 / Y < 2)$ (*vi*) $\rho_{X,Y}$.

Sol. $$f_X(x) = \int_0^\infty e^{-x-y} dy = e^{-x}\left\{-e^{-y}\right\}_0^\infty = e^{-x}$$

Similarly, $$f_Y(y) = e^{-y}$$

$\therefore$ $$f_{X,Y}(x,y) = e^{-x-y} = f_X(x)\, f_Y(y)$$

$\therefore$ X, Y are independent

$\therefore$ $$\rho_{X,Y} = 0$$

(*i*) $$P(X > 1) = \int_1^\infty f_X(x)\, dx = \int_1^\infty e^{-x} dx = e^{-1}$$

(V) $$P(0<X<1/Y<2) = P(0 < X < 1) \qquad \text{(as } X, Y \text{ are independent)}$$

$$= \int_0^1 e^{-x} dx = 1 - e^{-1}$$

(*ii*) To find dist. of $X + Y$, put

$$U = X + Y, \ V = X$$

Then $$X = V, Y = U - V$$

$\therefore$ $$0 \le V \le U \quad \text{and} \quad 0 \le U \le \infty$$

$$\frac{1}{J} = \begin{vmatrix} \frac{\partial U}{\partial X} & \frac{\partial U}{\partial Y} \\ \frac{\partial V}{\partial X} & \frac{\partial V}{\partial Y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

$\therefore$ Joint density function U, V i

$f_{U,V}(u,v)$

Marginal density of U is

$f_U(u)$

$\therefore$ $P(1 < X + Y < 2)$

(*iv*) $\frac{1}{2}$

$\therefore$ $(m+1)e^{-m}$

(*iii*) $P(X < Y \cap X < 2Y)$

$P(X < 2Y)$

$\cdot\, y+1)$

$\frac{y+1}{y+1)}$

$\frac{n-y+1)}{-2)}$

$\frac{n-y)!}{+1)!} = \frac{1}{n+1} \quad n = 0, 1, 2, \ldots\ldots n.$

$_{(0,\infty)}(y)$

$< Y|X < 2y)$

2) $(vi)\ \rho_{X,Y}$.

$\left\{-e^{-y}\right\}_0^{\infty} = e^{-y}$

$_Y(y)$

$^{-x}dx = e^{-1}$

(as X, Y are independent)

$= -1$

$\therefore$ Joint density function U, V is

$$f_{U,V}(u,v) = e^{-u}|J| = e^{-u}$$

Marginal density of U is

$$f_U(u) = \int_0^u e^{-u}dv = ue^{-u}$$

$$\therefore \quad P(1 < X+Y < 2) = P(1 < U < 2)$$

$$= \int_1^2 ue^{-u}du$$

$$= \left|-ue^{-u}\right|_1^2 + \int_1^2 e^{-u}du$$

$$= e^{-1} - 2e^{-2} + e^{-1} - e^{-2}$$

$$= 2e^{-1} - 3e^{-2}$$

(iv)

$$\frac{1}{2} = P(X+Y < m)$$

$$= P(U < m)$$

$$= \int_0^m ue^{-u}$$

$$= \left[-ue^{-u} - e^{-u}\right]_0^m = -(m+1)e^{-m} + 1$$

$$\therefore \quad (m+1)e^{-m} = \frac{1}{2} \quad \text{which gives } m.$$

(iii)

$$P(X < Y \cap X < 2Y) = P(X < Y)$$

$$= \int_0^{\infty} e^{-y}\left\{\int_0^y e^{-x}dy\right\}dy$$

$$= \int_0^{\infty} e^{-y}\left\{1 - e^{-y}\right\}dy$$

$$= 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(X < 2Y) = \int_0^{\infty} e^{-y}\left\{\int_0^{2y} e^{-x}dx\right\}dy$$

$$= \int_0^\infty e^{-y}\left\{1-e^{-2y}\right\}dy$$

$$= 1-\frac{1}{3}=\frac{2}{3}$$

$$\therefore \qquad P(X<Y|X<2Y) = \frac{1/2}{2/3} = 3/4.$$

Ex. 11-24. *For the bivariate distribution given by :*

$$f_{X,Y}(x,y) = e^{-y}(1-e^{-x})\, I_{(0,y)}(x)\, I_{(0,\infty)}(y) + e^{-x}(1-e^{-y})\, I_{(0,x)}(y)\, I_{(0,\infty)}(x)$$

(*i*) *Show that :* $f_{X,Y}(\cdot\cdot)$ *is p.d.f. and find*

(*ii*) *the marginal distributions of X and Y.*

(*iii*) $E\{Y|X=x\}$ *for* $0<x$

(*iv*) $P\{X\le 2; Y\le 2\}$

(*v*) ρ_{xy}

(*vi*) *another joint prob. density function having the same marginals.*

Sol. (*i*) $\displaystyle\int_x\int_y f_{X,Y}(x,y)\,dx\,dy = I_1+I_2$

where
$$I_1 = \int_0^\infty dy\left\{\int_0^y e^{-y}(1-e^{-x})dx\right\}$$

$$= \int_0^\infty e^{-y}dy\left\{y-1+e^{-y}\right\}$$

$$= \int_0^\infty ye^{-y}dy - \int_0^\infty e^{-y}dy + \int_0^\infty e^{-2y}dy$$

$$= 1-1+\frac{1}{2}=\frac{1}{2}$$

Similarly,
$$I_2 = \int_{x=0}^\infty dx\int_{y=0}^x e^{-x}(1-e^{-y})\,dy = \frac{1}{2}$$

$$\therefore \qquad \int\int f_{X,Y}(x,y)\,dx\,dy = \frac{1}{2}+\frac{1}{2}=1$$

Also $f_{X,Y}(x,y)\ge 0$

$\therefore$ $f_{X,Y}(\cdot\cdot)$ is prob. density functi

(*ii*) Marginal distributions are as l

$$f_Y(y) = \int_0^y e^{-y}(1-e^{-x})\,dx +$$

$$= ye^{-y}.$$

This is because : region of integra
region for given *y*, *x* varies from *y* to ∞

$$f_X(x) =$$

$$=$$

(*iii*) $$f_{Y/X}(y/x) =$$

$$=$$

$$= \frac{e^{-y}(1-e^{-x})\, I_{(x,\infty)}(y)\, I_{(0,\infty)}(x)\cdot}{xe^{-x}\, I}$$

$$= \frac{1}{xe^{-x}}\Big[e^{-y}(1-e^{-x})I_{(x,\infty)}(y)+e$$

$$\therefore\ E(Y|X) = \frac{1}{xe^{-x}}\Big[\int_x^\infty ye^{-y}(1-e$$

$$=1+\frac{x}{2}.$$

(*iv*) $$P\{X\le 2; Y\le 2\} =$$

dy

$- e^{-y})\, I_{(0,x)}(y)\, I_{(0,\infty)}(x)$

he same marginals.

$e^{-x})dx\Big\}$

$e^{-y}\Big\}$

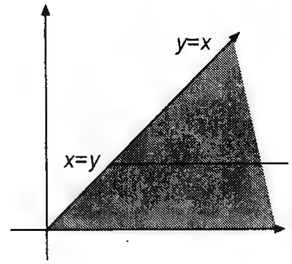
${}^{y}dy + \int_0^{\infty} e^{-2y}\,dy$

$e^{-y})\,dy = \frac{1}{2}$

Also $f_{X,Y}(x,y) \geq 0$

$\therefore$ $f_{X,Y}(\because)$ is prob. density function.

(*ii*) Marginal distributions are as below :



$$f_Y(y) = \int_0^y e^{-y}(1-e^{-x})\,dx + \int_y^{\infty} e^{-x}(1-e^{-y})\,dx$$

$$= ye^{-y}.$$

This is because : region of integration for second integral is shaded portion and in this region for given y, x varies from y to ∞.

$$f_X(x) = \int_x^{\infty} e^{-y}(1-e^{-x})\,dy + \int_0^x e^{-x}(1-e^{-y})\,dy$$

$$= xe^{-x}.$$

(*iii*)
$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \Big\{e^{-y}(1-e^{-x})\, I_{(0,y)}(x)\, I_{(0,\infty)}(y)$$

$$+ e^{-x}(1-e^{-y})\, I_{(0,x)}(y)\, I_{(0,\infty)}(x)\Big\}.\frac{1}{xe^{-x} I_{(0,\infty)}(x)}$$

$$= \frac{e^{-y}(1-e^{-x})\, I_{(x,\infty)}(y)\, I_{(0,\infty)}(x) + e^{-x}(1-e^{-y})\, I_{(0,x)}(y)\, I_{(0,\infty)}(x)}{xe^{-x}\, I_{(0,\infty)}(x)}$$

$$= \frac{1}{xe^{-x}}\Big[e^{-y}(1-e^{-x}) I_{(x,\infty)}(y) + e^{-x}(1-e^{-y})\, I_{(0,x)}(y)\Big]$$

$$\therefore\ E(Y|X) = \frac{1}{xe^{-x}}\left[\int_x^{\infty} ye^{-y}(1-e^{-x})\,dy + \int_0^x ye^{-x}(1-e^{-y})\,dy\right]$$

$$= 1 + \frac{x}{2}.$$

(*iv*)
$$P\{X \leq 2; Y \leq 2\} = \int_{y=0}^{2} dy \int_{x=0}^{y} e^{-y}(1-e^{-x})\,dx + \int_{x=0}^{2} dx \int_{y=0}^{x} e^{-x}(1-e^{-y})\,dy$$

$$= \int_{y=0}^{2} e^{-y}\left\{y + e^{-y} - 1\right\} dy + \int_{x=0}^{2} e^{-x}\left\{x + e^{-x} - 1\right\} dx$$

$$= 2\left\{\int_0^2 (x-1)e^{-x}\,dx + \int_0^2 e^{-2x}\,dx\right\}$$

$$= 2\left\{-xe^{-x} - \frac{e^{-2x}}{2}\right\}_0^2 = 1 - 4e^{-2} - e^{-4}$$

(v)
$$E(X) = \int_0^\infty x\, f_X(x) = \int_0^\infty x^2 e^{-x} dx = 2! = 2$$

$$E(X^2) = \int_0^\infty x^2 \cdot xe^{-x} dx = 3! = 6$$

$$\text{var}(X) = E(X^2) - \{E(X)\}^2 = 6 - 4 = 2$$

Similarly $\text{var}(Y) = 2$

$$\text{cov}(X,Y) = E(XY)$$

$$= \int_{y=0}^{\infty}\int_0^y xye^{-y}(1-e^{-x})dx\, dy + \int_{x=0}^{\infty}\int_{y=0}^{x} xye^{-x}(1-e^{-y})\, dx\, dy$$

$$= \int_{y=0}^{\infty} ye^{-y}\left\{\frac{y^2}{2} + (y+1)e^{-y} - 1\right\}dy$$

$$+\int_0^\infty xe^{-x}\left\{\frac{x^2}{2} + (x+1)e^{-x} - 1\right\}dx$$

$$= 2\int_0^\infty xe^{-x}\left\{\frac{x^2}{2} + (x+1)e^{-x} - 1\right\}dx$$

$$= 2\left[\frac{1}{2}\int_0^\infty x^3 e^{-x}dx + \int_0^\infty x^2 e^{-2x}dx + \int_0^\infty xe^{-2x}dx - \int_0^\infty xe^{-x}dx\right]$$

$$= 2\left[\frac{1}{2}(3!) + \frac{1}{4} + \frac{1}{4} - 1\right]$$

$$= 6 + 1 - 2 = 5$$

$$\text{cov}(X,Y) = E(XY) - E(X)\,E(Y)$$

$$= 5 - 4 = 1$$

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{(s.d \text{ of } X)(s.d \text{ of } Y)} = \frac{1}{2}$$

(vi) An infinite family of joint probability density functions, each having the marginals $f_X(x)$ and $f_Y(y)$ is

$$f_{X,Y}(x,y;\alpha) = f_X(x)\,f_Y(y)\left[1 + \alpha\{2F_X(x) - 1\}\{2F_Y(y) - 1\}\right]$$

where $-1 \le \alpha \le 1$ (See Ex: 11.7

Take α

∴ New joint probability densit

$f_{X,Y}(x,y)$

Ex. 11-25. *The joint probabilit*

$f_{X,Y}(x,y) = 3($

Find : (i) the marginal densit

(ii) $P(X+Y<0.5)$

(iii) $E\{Y/X=x\}, E(X),$

(iv) $\text{cov}(X,Y)$

Sol. The region of integration i

(i) For given x, y varies from 0 to $1-x$

$$\therefore\ f_X(x) = 3\int_0^{1-x}(x+y)\,dy$$

$$= 3\left[xy + \frac{y^2}{2}\right]_{y=0}^{1-x}$$

$$= \frac{3}{2}(1-x^2)$$

(ii) $P\{X+Y<0{\cdot}5\}$

$$= \int_{x=0}^{0\cdot5}\int_{y=0}^{0\cdot5-x} 3(x+y)\,dx$$

$$= 3\int_{x=0}^{\frac{1}{2}}\left(xy + \frac{y^2}{2}\right)_0^{\frac{1}{2}-}$$

$$= \frac{3}{2}\int_0^{\frac{1}{2}}\left(\frac{1}{4} - x^2\right)dx$$

$$= \frac{3}{2}\left\{\frac{1}{4}x - \frac{x^3}{3}\right\}_0^{1/2} =$$

$^{-2}-e^{-4}$

$=2$

2

$\int\limits_{x=0}^{\infty}\int\limits_{y=0}^{x} xye^{-x}(1-e^{-y})\,dx\,dy$

dy

$+\int_0^{\infty} xe^{-x}\left\{\frac{x^2}{2}+(x+1)e^{-x}-1\right\}dx$

$\Big\}dx$

$lx+\int_0^{\infty} xe^{-2x}dx-\int_0^{\infty} xe^{-x}dx\Big]$

nctions, each having the marginals

$\}\{2F_Y(y)-1\}]$

where $-1 \le \alpha \le 1$ (See Ex: 11.7).

Take $\alpha = 0$

$\therefore$ New joint probability density function is

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)\,f_Y(y) \\ &= (xe^{-x})(ye^{-y}) \\ &= xye^{-x-y}, \quad 0 \le x \le \infty \\ &\qquad\qquad\quad\ 0 \le y \le \infty. \end{aligned}$$

Ex. 11-25. *The joint probability density function of X and Y is as below :*

$$f_{X,Y}(x,y) = 3(x+y)I_{(0,1)}(x+y)I_{(0,1)}(x)I_{(0,1)}(y)$$

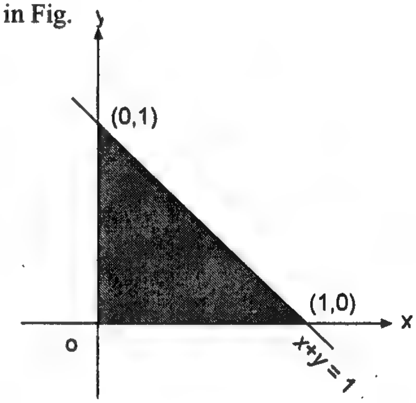
Find : (i) the marginal density of X

(ii) $P(X+Y<0.5)$

(iii) $E\{Y/X=x\}, E(X), \mathrm{var}(X), E(Y)$

(iv) $\mathrm{cov}(X,Y)$

Sol. The region of integration is shown shaded in Fig.



(i) For given x, y varies from

0 to $1-x$

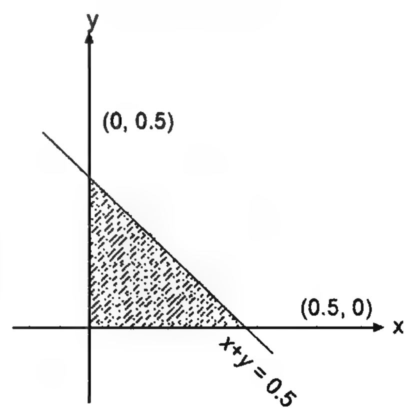
$$\therefore\ f_X(x) = 3\int_0^{1-x}(x+y)\,dy$$

$$= 3\left[xy+\frac{y^2}{2}\right]_{y=0}^{1-x}$$

$$= \frac{3}{2}(1-x^2)$$

(ii) $P\{X+Y<0\cdot5\}$

$$= \int\limits_{x=0}^{0\cdot5}\int\limits_{y=0}^{0\cdot5-x} 3(x+y)\,dx\,dy$$



$$= 3\int_{x=0}^{\frac{1}{2}}\left(xy+\frac{y^2}{2}\right)_0^{\frac{1}{2}-x}dx$$

$$= \frac{3}{2}\int_0^{\frac{1}{2}}\left(\frac{1}{4}-x^2\right)dx$$

$$= \frac{3}{2}\left\{\frac{1}{4}x-\frac{x^3}{3}\right\}_0^{1/2} = \frac{1}{8}$$

(*iii*)
$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x;y)}{f_X(x)} = \frac{3(x+y)}{\frac{3}{2}(1-x^2)}$$
$$= \frac{2(x+y)}{1-x^2}$$

$\therefore$
$$E\{Y/X=x\} = \int_0^{1-x} y.\frac{2(x+y)}{1-x^2}\,dy$$
$$= \frac{2}{1-x^2}\left\{x\frac{y^2}{2}+\frac{y^3}{3}\right\}_0^{1-x}$$
$$= \frac{1}{3(1-x^2)}\left\{x^3-3x+2\right\}$$

$\therefore$
$$E(Y) = E\left[E\{Y|X=x\}\right] = E\left[\frac{1}{3(1-x^2)}(x^3-3x+2)\right]$$
$$= \frac{1}{3}\int_0^1 \frac{1}{1-x^2}.(x^3-3x+2).\frac{3}{2}(1-x^2)\,dx$$
$$= \frac{1}{2}\int_0^1 (x^3-3x+2)\,dx = 3/8$$

(*iv*)
$$E(XY) = 3\int_{x=0}^1\left\{\int_{y=0}^{1-x} xy(x+y)\,dy\right\}dx$$
$$= \frac{1}{2}\int_0^1 x(x^3-3x+2)\,dx = \frac{1}{10}$$
$$E(X) = \int_0^1 x f_X(x)\,dx = \frac{3}{2}\int_0^1 x(1-x^2)\,dx = \frac{3}{8}$$
$$E(X^2) = \frac{3}{2}\int_0^1 x^2(1-x^2)\,dx = \frac{1}{5}$$
$$\text{var}(X) = E(X^2)-\{E(X)\}^2 = \frac{1}{5}-\frac{9}{64} = \frac{19}{320}$$
$$\text{cov}(X,Y) = E(XY)-E(X).E(Y)$$
$$= \frac{1}{10}-\frac{3}{8}.\frac{3}{8} = \frac{-13}{320}$$

Ex. 11-26. *The discrete density of X is given by*
$$f_X(x) = \frac{x}{3},\ x = 1, 2$$

and $f_{Y/X}(y/x)$ *is binomial with parameters x and* $\frac{1}{2}$.

i.e., $f_{Y/X}(y/x) = {}^xc_y\left(\frac{1}{2}\right)^x$, $y = 0, 1$

Find : (*i*) $E(X)$, var(X)

(*ii*) $E(Y)$

(*iii*) *Joint distribution o*

Sol. (*i*) $E(X)$

$E(X^2)$

$\therefore$ var $(X$

(*ii*) $E(Y/X$

$E(Y$

(*iii*) $f_{X,Y}(x,y$

11.6. Bivariate Normal Distribut

Let x and y be linearly correla same.

Now Eq. of line of regression

$y -$

$\Rightarrow$

Where ρ = Correlation co-effi

Which gives the mean of the

Also standard error of estimat

This is also the s.d. of y for gi

$\therefore$ Conditional density f^n of

$\sigma_y\sqrt{1-}$

i.e., $f_{Y/X}(y/x) = {}^{x}c_y \left(\frac{1}{2}\right)^x$, $y = 0, 1, 2 \ldots\ldots x.$

Find : (*i*) $E(X)$, $\text{var}(X)$

(*ii*) $E(Y)$

(*iii*) *Joint distribution of X and Y.*

Sol. (*i*)

$$E(X) = \Sigma\, x \cdot \frac{x}{3} = \frac{1}{3} + \frac{4}{3} = \frac{5}{3}$$

$$E(X^2) = \Sigma\, x^2 \cdot \frac{x}{3} = \frac{1}{3} + \frac{8}{3} = 3$$

$$\therefore \quad \text{var}(X) = E(X^2) - (E(X))^2 = 3 - \frac{25}{9} = \frac{2}{9}$$

(*ii*)

$$E(Y/X) = x \cdot \frac{1}{2} = \frac{x}{2}$$

$$E(Y) = E\{E(Y/X)\} = E\left(\frac{x}{2}\right) = \frac{E(x)}{2} = \frac{5}{6}$$

(*iii*)

$$f_{X,Y}(x,y) = f_{Y/X}(y/x) \cdot f_X(x)$$

$$= {}^{x}c_y \cdot \left(\frac{1}{2}\right)^x \cdot \frac{x}{3}.$$

11.6. Bivariate Normal Distribution

Let x and y be linearly correlated normal variates such that variance of y for each x is same.

Now Eq. of line of regression of y on x is

$$y - \bar{y} = \rho \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$\Rightarrow \quad y = \bar{y} + \rho \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

Where ρ = Correlation co-eff. between x, y.

Which gives the mean of the values of y for given x.

Also standard error of estimate of y on x is $\sigma_y \sqrt{1-\rho^2}$

This is also the s.d. of y for given x.

$\therefore$ Conditional density f^n of y for given x is

$$\frac{1}{\sigma_y \sqrt{1-\rho^2}\,\sqrt{2\pi}}\, e^{-\frac{1}{2}\left\{\frac{y - \left\{\left(\bar{y} + \rho\frac{\sigma_y}{\sigma_x}(x-\bar{x})\right)\right\}}{\sigma_y\sqrt{1-\rho^2}}\right\}^2}$$

$$\frac{1}{\sigma_y\sqrt{1-\rho^2}\,\sqrt{2\pi}}\,e^{-\frac{1}{2}\left\{\frac{(y-\bar{y})-\rho\frac{\sigma_y}{\sigma_x}(x-\bar{x})}{\sigma_y\sqrt{1-\rho^2}}\right\}^2}$$

Now the density f^n of x is

$$\frac{1}{\sigma_x\sqrt{2\pi}}\,e^{-\frac{1}{2}\left\{\frac{x-\bar{x}}{\sigma_x}\right\}^2}$$

∴ Prob. density f^n is given by

$$f(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma_x.\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\,e^{-\frac{1}{2}\left[\left\{\frac{x-\bar{x}}{\sigma_x}\right\}^2+\left\{\frac{(y-\bar{y})-\rho\frac{\sigma_y}{\sigma_x}(x-\bar{x})}{\sigma_y\sqrt{1-\rho^2}}\right\}^2\right]}$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\bar{x}}{\sigma_x}\right)\left(\frac{y-\bar{y}}{\sigma_y}\right)+\left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}\right] \quad \ldots(1)$$

Both x and y vary from $-\infty$ to ∞

$$\therefore \text{Total prob.} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\bar{x}}{\sigma_x}\right)\left(\frac{y-\bar{y}}{\sigma_y}\right)+\left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}\right]dx\,dy$$

Put $X = \frac{x-x}{\sigma_x}, Y = \frac{y-y}{\sigma_y}$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2(1-\rho^2)}\{X^2-2\rho XY+Y^2\}\right]dX\,dY$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2(1-\rho^2)}\{(X-\rho Y)^2+(1-\rho^2)Y^2\}\right]dX\,dY$$

For def. of correlation see chapter 13.

Put $\frac{X-\rho Y}{\sqrt{1-\rho^2}} = u, Y =$

$$\therefore \frac{\partial(u,v)}{\partial(X,Y)} =$$

Total prob. $= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+}$

∴ (1) Represents a density f^n.

normal distribution.

11.6.1. Marginal and Conditiona

Marginal density of x is given

$$f_x(x) = \int_{-\infty}^{\infty} f(x,$$

Put $\frac{y-\bar{y}}{\sigma_y}$

Put $$\frac{X-\rho Y}{\sqrt{1-\rho^2}} = u, Y = v$$

$$\therefore \quad \frac{\partial(u,v)}{\partial(X,Y)} = \begin{vmatrix} \dfrac{\partial u}{\partial X} & \dfrac{\partial v}{\partial X} \\[2ex] \dfrac{\partial u}{\partial Y} & \dfrac{\partial v}{\partial Y} \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{1}{\sqrt{1-\rho^2}} & 0 \\[2ex] -\dfrac{\rho}{\sqrt{1-\rho^2}} & 1 \end{vmatrix} = \frac{1}{\sqrt{1-\rho^2}}$$

$$\text{Total prob.} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+v^2)}\,du\,dv = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-u^2}\,du.\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-v^2}\,dv = 1$$

$\therefore$ (1) Represents a density f^n. The bivariate distribution given by (1) is called **Bivariate normal distribution.**

11.6.1. Marginal and Conditional Densities

Marginal density of x is given by

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2(1-\rho)^2}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\bar{x}}{\sigma_x}\right)\left(\frac{y-\bar{y}}{\sigma_y}\right)+\left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}\right]dy$$

Put $$\frac{y-\bar{y}}{\sigma_y} = Y, \qquad \frac{x-\bar{x}}{\sigma_x} = X$$

$$= \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left[\frac{X^2+Y^2-2\rho XY}{-2(1-\rho^2)}\right]dY$$

$$= \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}}\,e^{-\frac{1}{2}X^2}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2(1-\rho^2)}{(Y-\rho X)^2}\right]dY$$

$$= \frac{1}{2\pi\sigma_x} e^{-\frac{1}{2}X^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\vartheta^2} d\vartheta. \qquad \text{where } \vartheta = \frac{Y-\rho X}{\sqrt{1-\rho^2}}$$

$$= \frac{1}{\sqrt{2\pi}.\sigma_x} e^{-\frac{1}{2}X^2} . \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{-\frac{1}{2}\vartheta^2} d\vartheta$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma_x}\right)^2}$$

Similarly marginal density of y is given by

$$f_y(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}}{\sigma_y}\right)^2}$$

Conditional density f^n of x given y is given by

$$f_{x/y}(x/y) = \frac{f(x,y)}{f_y(y)}$$

$$= \frac{1}{\sqrt{2\pi}.\sigma_x.\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{x-\bar{x}}{\sigma_x} - \rho\frac{y-\bar{y}}{\sigma_y}\right\}^2\right]$$

$$= \frac{1}{\sqrt{2\pi}.\sigma_x.\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\left\{\frac{x-\left(\bar{x}+\rho\frac{\sigma_x}{\sigma_y}(y-\bar{y})\right)}{\sigma_x\sqrt{1-\rho^2}}\right\}^2\right]$$

$\Rightarrow$ Conditional distribution of x given y is normal with mean $\bar{x}+\rho\dfrac{\sigma_x}{\sigma_y}(y-\bar{y})$ and s.d. $\sigma_x\sqrt{1-\rho^2}$

Similarly condition distribution of y given x is normal with mean $\bar{y}+\rho\dfrac{\sigma_y}{\sigma_x}(x-\bar{x})$ and s.d. $\sigma_y\sqrt{1-\rho^2}$.

11.6.2. Moment Generating Function

$$M_{0,0}(t_1,t_2) = E\left\{e^{t_1x+t_2y}\right\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+t_2y} f(x,y)\, dx\, dy$$

Put $X = \frac{x-\bar{x}}{\sigma_x}$,

$\Rightarrow$ $dx = \sigma_x dX$,

$\therefore$ $M_{0,0}(t_1,t_2) = \frac{e^{t_1\bar{x}+}}{2\pi\sqrt{1-}}$

$= \frac{e^{t_1\bar{x}+}}{2\pi\sqrt{1-}}$

Then term within { } brack

$X^2-2\rho XY+Y^2-2(1-\rho^2)($

$= (X-\rho Y)^2-2(1-\rho^2)t_1\sigma_x($

$= \{X-\rho Y-(1-\rho^2)t_1\sigma_x\}^2+$

$= \{X-\rho Y-(1-\rho^2)t_1\sigma_x\}^2+$

$= \{X-\rho Y-(1-\rho^2)t_1\sigma_x\}^2+$

$\therefore$ $M_{0,0}(t_1,t_2) = e^{t_1\bar{x}+t_2}$

$\frac{1}{2\pi\sqrt{1-\rho^2}}\int\int$ ex

Put X

$\frac{\partial(u,v)}{\partial(X,Y)} = \begin{vmatrix} \frac{\partial u}{\partial X} \\ \frac{\partial v}{\partial X} \end{vmatrix}$

where $\vartheta = \frac{Y-\rho X}{\sqrt{1-\rho^2}}$

ϑ^2 $d\vartheta$

$$\frac{1}{-\rho^2)}\left\{\frac{x-\bar{x}}{\sigma_x}-\rho\frac{y-\bar{y}}{\sigma_y}\right\}^2\Bigg]$$

$$\left.\frac{x-\left(\bar{x}+\rho\frac{\sigma_x}{\sigma_y}(y-\bar{y})\right)}{\sigma_x\sqrt{1-\rho^2}}\right\}^2\Bigg]$$

vith mean $\bar{x}+\rho\frac{\sigma_x}{\sigma_y}(y-\bar{y})$ and s.d.

ıal with mean $\bar{y}+\rho\frac{\sigma_y}{\sigma_x}(x-\bar{x})$ and

Put $$X = \frac{x-\bar{x}}{\sigma_x},\ Y=\frac{y-\bar{y}}{\sigma_y}$$

$$\Rightarrow \qquad dx = \sigma_x dX,\ dy = \sigma_y dY$$

$$\therefore \quad M_{0,0}(t_1,t_2) = \frac{e^{t_1\bar{x}+t_2\bar{y}}}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1\sigma_x X+t_2\sigma_y Y}\, e^{-\frac{1}{2(1-\rho^2)}(X^2-2\rho XY+Y^2)}\,dX\,dY$$

$$= \frac{e^{t_1\bar{x}+t_2\bar{y}}}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[\frac{-1}{2(1-\rho^2)}\{X^2-2\rho XY+Y^2\right.$$

$$\left.-2(1-\rho^2)(t_1\sigma_x X+t_2\sigma_y Y)\}\right]dX\,dY$$

Then term within { } bracket is

$$X^2-2\rho XY+Y^2-2(1-\rho^2)(t_1\sigma_x X+t_2\sigma_y Y)$$

$$=(X-\rho Y)^2-2(1-\rho^2)t_1\sigma_x(X-\rho Y)+(1-\rho^2)\{Y^2-2\rho t_1\sigma_x Y-2t_2\sigma_y Y\}$$

$$=\{X-\rho Y-(1-\rho^2)t_1\sigma_x\}^2+(1-\rho^2)\{Y^2-2\rho t_1\sigma_x Y-(1-\rho^2)t_1^2\sigma_x^2-2t_2\sigma_y Y\}$$

$$=\{X-\rho Y-(1-\rho^2)t_1\sigma_x\}^2+(1-\rho^2)\{(Y-\rho t_1\sigma_x)^2$$

$$-2t_2\sigma_y(Y-\rho t_1\sigma_x)-t_1^2\sigma_x^2-2\rho t_1t_2\sigma_x\sigma_y\}$$

$$=\{X-\rho Y-(1-\rho^2)t_1\sigma_x\}^2+(1-\rho^2)\{Y-\rho t_1\sigma_x-t_2\sigma_y)^2$$

$$-(t_1\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\}$$

$$\therefore \quad M_{0,0}(t_1,t_2) = e^{t_1\bar{x}+t_2\bar{y}+\frac{1}{2}(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)}$$

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\iint\exp\left\{\frac{\begin{array}{c}(X-\rho Y-(1-\rho^2)t_1\sigma_x)^2\\+(1-\rho^2)(Y-\rho t_1\sigma_x-t_2\sigma_y)^2\end{array}}{-2(1-\rho^2)}\right\}dX\,dY$$

Put $$X-\rho Y-(1-\rho^2)t_1\sigma_x = u\sqrt{1-\rho^2}$$

$$Y-\rho t_1\sigma_x-t_2\sigma_y = v$$

$$\frac{\partial(u,v)}{\partial(X,Y)} = \begin{vmatrix}\frac{\partial u}{\partial X} & \frac{\partial u}{\partial Y}\\ \frac{\partial v}{\partial X} & \frac{\partial v}{\partial Y}\end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{1}{\sqrt{1-\rho^2}} & \dfrac{-\rho}{\sqrt{1-\rho^2}} \\ 0 & 1 \end{vmatrix} = \frac{1}{\sqrt{1-\rho^2}}$$

$\therefore \; M_{0,0}(t_1, t_2)$

$$= \exp\left[t_1\bar{x} + t_2\bar{y} + \frac{1}{2}(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x\sigma_y + t_2^2\sigma_y^2\right] \cdot \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+v^2)}\,du\,dv$$

$$= \exp\left[t_1\bar{x} + t_2\bar{y} + \frac{1}{2}\left(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x\sigma_y + t_2^2\sigma_y^2\right)\right] \left\{\frac{1}{\sqrt{2\pi}}\int_{-\infty} e^{-\frac{1}{2}u^2}du\right\}\left\{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2}dv\right\}$$

$$= \exp\left[t_1\bar{x} + t_2\bar{y} + \frac{1}{2}\left(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x\sigma_y + t_2^2\sigma_y^2\right)\right]$$

Cor. (1) if $\bar{x} = \bar{y} = 0, \qquad \sigma_x = \sigma_y = 1$

$$M_{0,0}(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)}$$

(2) $$M_{\bar{x},\bar{y}}(t_1, t_2) = E\{e^{t_1(x-\bar{x}) + t_2(y-\bar{y})}\}$$

$$= e^{-(t_1\bar{x} + t_2\bar{y})} M_{0,0}(t_1, t_2)$$

$$= e^{\frac{1}{2}(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x\sigma_y + t_2^2\sigma_y^2)}$$

which gives *m.g.f.* about $(\bar{x}, \bar{y})$

Ex. 11-27. *For the bivariate normal distribution :*

$$dP = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(x^2 - 2\rho xy + y^2\right)\right]$$

where x, y denote the deviations from their means, find *m.g.f.* about mean $(\bar{x}, \bar{y})$. Deduce that

$\mu_{rs} = (r+s-1)\rho\mu_{r-1,s-1} + (r-1)(s-1)(1-\rho^2)\mu_{r-2,s-2}$ *and hence show that*

$\mu_{rs} = 0$ if $(r+s)$ is odd.

Sol. As in, cor. 1 it can be shown

$$M_{\bar{x},\bar{y}}(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + 2\ldots}$$

let $M_{\bar{x},\bar{y}}(t_1, t_2)$ be denot

$\therefore$ $$M = e^{\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)}$$

$\Rightarrow$ $$\log M = \frac{1}{2}(t_1^2 + 2\rho t_1 t\ldots$$

Differentiate partially

$\therefore$ $$\frac{1}{M}\frac{\partial M}{\partial t_1} = t_1 + \rho t_2$$

$\Rightarrow$ $$\frac{\partial M}{\partial t_1} = M(t_1 + \rho t_2)$$

and $$\frac{\partial M}{\partial t_2} = M(t_2 + \rho t_1)$$

Differentiating (3) *w.r.t.* t_1 partial

$$\frac{\partial^2 M}{\partial t_1 \partial t_2} = \frac{\partial M}{\partial t_1}(t_2 + \rho\ldots$$

$$= M\{t_1 + \rho t_2\ldots$$

$$= M\{t_1 t_2(1+\ldots$$

$$\therefore \frac{\partial^2 M}{\partial t_1 \partial t_2} - \rho t_1 \frac{\partial M}{\partial t_1} - \rho t_2 \frac{\partial M}{\partial t_2}$$

$$= M\left\{t_1 t_2(1+\rho^2) + \rho(t_1^2 + t_2^2 + 1) - \ldots\right.$$

$$= M\left\{\rho + (1-\rho^2)t_1 t_2\right\}$$

Now $$M = \sum_r\sum_s \mu_{rs}\ldots$$

$\therefore$ $$\frac{\partial M}{\partial t_1} = \sum_r\sum_s \mu_{rs}\ldots$$

$$\frac{\partial M}{\partial t_2} = \sum_r\sum_s \mu_{rs}\ldots$$

$$\frac{\partial^2 M}{\partial t_1 \partial t_2} = \sum_r\sum_s \mu_{rs}\ldots$$

$$= \frac{1}{\sqrt{1-\rho^2}}$$

$+ 2\rho t_1 t_2\, \sigma_x\sigma_y + t_2^2\sigma_y^2 \Big]$

$e^{-\frac{1}{2}(u^2+v^2)}\, du\, dv$

$2\rho t_1 t_2\, \sigma_x\sigma_y + t_2^2\,\sigma_y^2\big)\Big]$

$\frac{1}{\sqrt{2\pi}} \int_{-\infty} e^{-\frac{1}{2}u^2} du\Big\}\Big\{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2} dv\Big\}$

$2\rho t_1 t_2\sigma_x\sigma_y + t_2^2\sigma_y^2\big)\Big]$

$_x = \sigma_y = 1$

$\left(x^2 - 2\rho xy + y^2\right)\Big]$

d *m.g.f.* about mean $(\bar{x}, \bar{y})$. Deduce

$_{-2,s-2}$ *and hence show that*

Sol. As in, cor. 1 it can be shown that

$$M_{\bar{x},\bar{y}}(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)}$$

let $\quad M_{\bar{x},\bar{y}}(t_1, t_2)$ be denoted by M.

$$\therefore \quad M = e^{\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)}$$

$$\Rightarrow \quad \log M = \frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2) \qquad \ldots(1)$$

Differentiate partially *w.r.t.* t_1, t_2 respectively

$$\therefore \quad \frac{1}{M}\frac{\partial M}{\partial t_1} = t_1 + \rho t_2$$

$$\Rightarrow \quad \frac{\partial M}{\partial t_1} = M(t_1 + \rho t_2) \qquad \ldots(2)$$

and $$\frac{\partial M}{\partial t_2} = M(t_2 + \rho t_1) \qquad \ldots(3)$$

Differentiating (3) *w.r.t.* t_1 partially.

$$\frac{\partial^2 M}{\partial t_1 \partial t_2} = \frac{\partial M}{\partial t_1}(t_2 + \rho t_1) + M\rho$$

$$= M\{t_1 + \rho t_2)(t_2 + \rho t_1) + \rho\}$$

$$= M\{t_1 t_2(1+\rho^2) + \rho(t_1^2 + t_2^2 + 1)\}$$

$$\therefore \quad \frac{\partial^2 M}{\partial t_1 \partial t_2} - \rho t_1 \frac{\partial M}{\partial t_1} - \rho t_2 \frac{\partial M}{\partial t_2}$$

$$= M\left\{t_1 t_2(1+\rho^2) + \rho(t_1^2 + t_2^2 + 1) - \rho t_1(t_1 + \rho t_2) - \rho t_2(t_2 + \rho t_1)\right\}$$

$$= M\left\{\rho + (1-\rho^2)t_1 t_2\right\}$$

Now $$M = \sum_r \sum_s \mu_{rs} \frac{t_1^r}{r!}\frac{t_2^s}{s!}$$

$$\therefore \quad \frac{\partial M}{\partial t_1} = \sum_r \sum_s \mu_{rs} \frac{t_1^{r-1}}{(r-1)!}\frac{t_2^s}{s!}$$

$$\frac{\partial M}{\partial t_2} = \sum_r \sum_s \mu_{rs} \frac{t_1^r}{r!}\frac{t_2^{s-1}}{(s-1)!}$$

$$\frac{\partial^2 M}{\partial t_1 \partial t_2} = \sum_r \sum_s \mu_{rs} \frac{t_1^{r-1}}{(r-1)!}\frac{t_2^{s-1}}{(s-1)!}$$

Substituting in (4)

$$\sum_r\sum_s \mu_{rs}\frac{t_1^{r-1}}{(r-1)!}\frac{t_2^{s-1}}{(s-1)!} - \rho\sum_r\sum_s\mu_{rs}\frac{t_1^r}{(r-1)!}\frac{t_2^s}{s!} - \rho\sum_r\sum_s\mu_{rs}\frac{t_1^r}{r!}\frac{t_2^s}{(s-1)!}$$

$$= \rho\sum_r\sum_s \mu_{rs}\frac{t_1^r}{r!}\frac{t_2^s}{s!} + (1-\rho^2)\sum_r\sum_s\mu_{rs}\frac{t_1^{r+1}}{r!}\frac{t_2^{s+1}}{s!}$$

Equating Co-efficients of $\dfrac{t_1^{r-1}}{(r-1)!}\dfrac{t_2^{s-1}}{(s-1)!}$

$$\mu_{rs} - \rho(r-1)\mu_{r-1,s-1} - \rho(s-1)\mu_{r-1,s-1}$$

$$= \rho\mu_{r-1,s-1} + (1-\rho^2)(r-1)(s-1)\mu_{r-2,s-2}$$

$$\Rightarrow \mu_{rs} = (r+s-1)\rho\mu_{r-1,s-1} + (r-1)(s-1)(1-\rho^2)\mu_{r-2,s-2} \qquad \ldots(5)$$

we have $\mu_{01} = \mu_{10} = 0$

Put $r = 1, s = 2$

$$\mu_{12} = (1+2-1)\rho\mu_{01} + 0 = 0$$

Similarly $\mu_{21} = 0.$

Put $r = 2, s = 3$

$$\mu_{23} = (2+3-1)\rho\mu_{12} + 1\cdot 2.(1-\rho^2)\mu_{0,1}$$
$$= 0.$$

Similarly $\mu_{32} = 0.$

Also we have $\mu_{ro} = \mu_{os} = 0$ if r, s are odd.

Now if R + s is odd, so are

$\therefore \quad (r-1)+(s-1) = r+s-2$

and $\quad (r-2)+(s-2) = r+s-4$

and so on

$\therefore$ We have

$$\mu_{rs} = 0 \text{ if } (r+s) \text{ is odd.}$$

Theorem 11.6.3. If (x, y) have a bivariate normal distribution then x and y are independent iff x and y are uncorrelated.

Proof. If x and y are independent then

Cor $(x, y) = 0$ *i.e.*, x and y are uncorrelated.

Converse. Let x and y be uncorrelated

$$\Rightarrow \rho = 0.$$

$$\therefore \quad f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left\{\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}\right]$$

$= \dfrac{1}{\sigma_x\sqrt{2\pi}}$

$= f_x(x).f_y$

$\Rightarrow$ x and y

Remark : If (x, y) don't have

that x, y are independent whenever

and y are normal (but their joint dist.

Ex. 11-28. For the bivariate nor

dP = const exp

show that correlation coefficient bet

Sol. Here

$\bar{x}_1 = 0,$

$\therefore$ $E(x$

$\Rightarrow$

Similarly

Now joint m.g.f. of x_1, x_2 is

$M_{0,0}(t_1, t_2) = e^{\frac{1}{2}(t_1^2+t_2^2+2}$

$= 1+\frac{1}{2}(t_1^2 +$

$\therefore \quad \dfrac{E(x_1^4)}{4!} = \dfrac{\mu_4'(0) \text{ of}}{4!}$

= Co-eff of

$= \dfrac{1}{8}$

$\therefore \quad E(x_1^4) = 3.$

$\therefore \; E(x_1^4) = 3$

Similarly $E(x_2^4) = 3$

Also $\dfrac{E(x_1^2x_2^2)}{2!2!} = \dfrac{\mu'_{22}(0,0)}{2!2!}$

= Co-eff of

$$= \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\left(\frac{x-\bar{x}}{\sigma_y}\right)^2\right\} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-\bar{y}}{\sigma_y}\right)^2\right\}\right]$$

$$= f_x(x).f_y(y).$$

$\Rightarrow$ x and y are independent.

Remark : If (x, y) don't have a bivariate normal distribution then it is not necessary that x, y are independent whenever $\rho = 0$. This is so even when marginal distributions of x and y are normal (but their joint dist. is not bivariate normal.)

Ex. 11-28. For the bivariate normal distribution

$$dP = \text{const} \exp\left\{\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{-2(1-\rho^2)}\right\}$$

show that correlation coefficient between $x_1{}^2$ and $x_2{}^2$ is ρ^2

Sol. Here

$$\bar{x}_1 = 0, \qquad \text{var}(x_1) = 1$$

$$\therefore \qquad E(x_1^2) - \bar{x}_1^2 = 1$$

$$\Rightarrow \qquad E(x_1^2) = 1$$

Similarly $\qquad E(x_2{}^2) = 1$

Now joint m.g.f. of x_1, x_2 is

$$M_{0,0}(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + t_2^2 + 2\rho t_1 t_2)}$$

$$= 1 + \frac{1}{2}\left(t_1^2 + t_2^2 + 2\rho t_1 t_2\right) + \frac{1}{2!}\left\{\frac{1}{2}\left(t_1^2 + t_2^2 + 2\rho t_1 t_2\right)^2\right\}^2 + \ldots\ldots$$

$$\therefore \qquad \frac{E(x_1^4)}{4!} = \frac{\mu_4{}'(0) \text{ of } x_1}{4!}$$

$$= \text{Co-eff of } t_1^4$$

$$= \frac{1}{8}$$

$$\therefore \qquad E(x_1{}^4) = 3.$$

$\therefore\ E(x_1{}^4) = 3$

Similarly $\quad E(x_2{}^4) = 3$

Also $\quad \dfrac{E(x_1^2 x_2^2)}{2!2!} = \dfrac{\mu'_{22}(0,0)}{2!2!}$

$$= \text{Co-eff of } t_1^2 t_2^2$$

$$= \frac{1}{2!}\left(\rho^2 + \frac{1}{2}\right)$$

$$\therefore \quad E(x_1{}^2 x_2{}^2) = 2\rho^2 + 1$$

$$\therefore \quad \gamma_{x_1^2, x_2^2} = \frac{\text{cov}\,(x_1^2, x_2^2)}{(\text{s.d. of } x_1^2)\,(\text{s.d. of } x_2^2)}$$

$$= \frac{E(x_1^2 x_2^2) - E(x_1^2)E(x_2^2)}{\sqrt{\left\{E(x_1^4) - (E(x_1^2))^2\right\} \cdot \left\{E(x_2^4) - (E(x_2^2))^2\right\}}}$$

$$= \frac{2\rho^2 + 1 - 1}{\sqrt{(3-1)(3-1)}} = \rho^2.$$

Ex. 11-29. *Let X and Y have bivariate normal distribution with parameters* $\bar{x} = 5, \sigma_x = 1$, $\bar{y} = 10$ and $\sigma_y = 5$.

(*i*) *if* $\rho > 0$, *find* ρ *such that*

$$P(4 < y < 16 | x = 5) = 0.954$$

(*ii*) *if* $\rho = 0$, *find* $P(x + y \le 16)$.

Sol.

$$f_{x,y}(x, y) = \frac{1}{2\pi \cdot 5\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-5}{1}\right)^2 - 2\rho\left(\frac{x-5}{1}\right)\left(\frac{y-10}{5}\right) + \left(\frac{y-10}{5}\right)^2\right\}\right]$$

and $$f_x(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-5}{1}\right)^2\right]$$

$$\therefore \quad f_{y/x}(y/x) = \frac{f_{x,y}(x,y)}{f_x(x)}$$

$$= \frac{1}{\sqrt{2\pi} \cdot 5\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{y-10}{5} - \rho\left(\frac{x-5}{1}\right)\right\}^2\right]$$

$$\therefore \quad f_{y/x}(y | x = 5) = \frac{1}{\sqrt{2\pi} \cdot 5\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{y-10}{5}\right)^2\right]$$

$$\therefore \quad E(y / x = 5) = 10$$

and $$\text{var}\,(y / x = 5) = \left(5\sqrt{1-\rho^2}\right)^2$$

Put $$z = \frac{y-10}{5\sqrt{1-\rho^2}}$$

$$\therefore \quad P\{4 < y < 16 | x = 5\}$$

$$\therefore \quad P\left\{0 < z < \frac{6}{5\sqrt{1-\rho^2}} \middle| x = 5\right\}$$

$\therefore$ From Normal Tables.

$$\frac{6}{5\sqrt{1-\rho^2}}$$

$$\therefore \quad 1 - \rho^2$$

(*ii*) If $\rho = 0$, X and Y are indep

x is a $N(5, 1)$

y is a $N(10, 5)$

$\therefore$ $t = x + y$ is a $N(15, \sqrt{26})$

$$P(x + y \le 16)$$

Put u

11.7. Bivariate Transformation

Let x and y be two random vari

Let x and y be transformed to variate

x

where u, v are continuously different

Let J

Then the joint probability densit

$|J$

Ex. 11-30. *If the probability den*

$f(x, y)$

$$\therefore \quad P\{4 < y < 16 | x = 5\} = P\left\{\frac{-6}{5\sqrt{1-\rho^2}} < z < \frac{6}{5\sqrt{1-\rho^2}} \Big| x = 5\right\}$$

$$= 2P\left\{0 < z < \frac{6}{5\sqrt{1-\rho^2}} \Big| x = 5\right\}$$

$$\therefore \quad P\left\{0 < z < \frac{6}{5\sqrt{1-\rho^2}} \Big| x = 5\right\} = \frac{0\cdot 954}{2} = 0\cdot 477$$

$\therefore$ From Normal Tables.

$$\frac{6}{5\sqrt{1-\rho^2}} = 2$$

$$\therefore \quad 1-\rho^2 = 0\cdot 36 \Rightarrow \rho = 0\cdot 8$$

(*ii*) If $\rho = 0$, X and Y are independent

x is a $N(5, 1)$

y is a $N(10, 5)$

$\therefore$ $t = x + y$ is a $N(15, \sqrt{26})$

$$P(x + y \le 16) = P(t \le 16)$$

Put
$$u = \frac{t-15}{\sqrt{26}}$$

$$= P\left(u < \frac{1}{\sqrt{26}} = 0\cdot 196\right)$$

$$= 0\cdot 5793 \text{ (using normal tables).}$$

11.7. Bivariate Transformation

Let x and y be two random variates having joint probability density function $f(x, y)$. Let x and y be transformed to variates u and v by the transformation

$$x = x(u, v);\ y = y(u, v)$$

where u, v are continuously differentiable function.

Let
$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\[2ex] \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \neq 0.$$

Then the joint probability density function of u and v is

$$|J|\ f\{x(u, v), y(u, v)\}.$$

Ex. 11-30. *If the probability density function of two variates x and y is given by*

$$f(x, y) = \begin{cases} 4xy\ e^{-(x^2+y^2)}; & x \ge 0,\quad y \ge 0 \\ 0 & \textit{elsewhere.} \end{cases}$$

Find the density function of $\sqrt{x^2+y^2}$.

Sol. Put $\qquad u = \sqrt{x^2+y^2}, \quad v = x$

Then $\qquad v \ge 0, \quad u \ge v.$

∴ $\qquad u \ge 0, \quad 0 \le v \le u.$

Now
$$\frac{1}{J} = \frac{\partial(u,v)}{\partial(x,y)}$$

$$= \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ 1 & 0 \end{vmatrix} = \frac{-y}{\sqrt{x^2+y^2}}$$

∴ The probability density function of u and v is given by

$$f_{u,v}(u,v) = f(x,y)|J|$$

$$= \begin{cases} \left(4xye^{-(x^2+y^2)}\right)\dfrac{\sqrt{x^2+y^2}}{y}, & x \ge 0, y \ge 0 \\ 0 \textit{ elsewhere} \end{cases}$$

$$= \begin{cases} 4uve^{-u^2} & u \ge 0, 0 \le v \le u. \\ 0 & \text{elsewhere} \end{cases}$$

∴ Marginal density function of u is

$$4\int_{v=0}^{u} uve^{-u^2}\, dv = 2u^3 e^{-u^2}, u \ge 0$$

∴ Density function of u is

$$\begin{cases} 2u^3e^{-u^2} & ,u \ge 0 \\ 0, & \text{elsewhere} \end{cases}$$

Ex. 11-31. *The joint density function of x and y is*

$$f(x,y) = e^{-(x+y)}, x > 0, y > 0.$$

Find the prob. density function of $\dfrac{x+y}{2}$.

Sol. Put $\qquad u = \dfrac{x+y}{2}, v = x.$

Then $v \ge 0, \ 0 \le v \le 2u$

$\frac{1}{J} =$

$=$

∴ The prob. density function of u

$=$

∴ Density function of u is

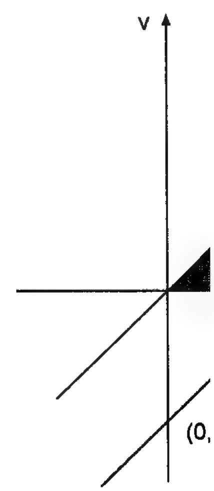
$=$

Ex. 11-32. *The joint density funct*

$f(x,y) =$

$=$

Find the distribution of x + y.



Sol. Put $u = x+y, v = y$

∴ $\quad x = u - v, y = v$

$y > 0 \Rightarrow v > 0$

$0 < x < 1 \Rightarrow 0 < u - v < 1$

x

$\frac{y}{{}^2+y^2} \Big| = \frac{-y}{\sqrt{x^2+y^2}}$
0

ven by

$\frac{\sqrt{x^2+y^2}}{y}, x \geq 0, y \geq 0$

$u \geq 0, 0 \leq v \leq u.$
elsewhere

$u \geq 0$

re

$> 0.$

Then $v \geq 0,\ 0 \leq v \leq 2u$

$$\frac{1}{J} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{vmatrix} = -\frac{1}{2}$$

$\therefore$ The prob. density function of u and v is

$$e^{-(x+y)}.|J|$$

$$= 2e^{-2u}$$

$\therefore$ Density function of u is

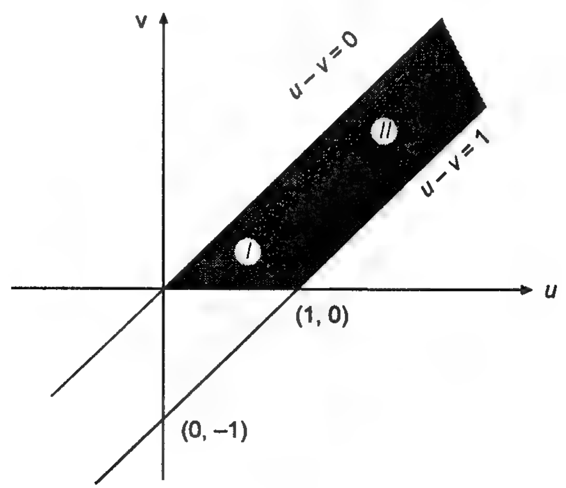
$$2\int_0^{2u} e^{-2u}\,dv$$

$$= 4ue^{-2u}.$$

Ex. 11-32. *The joint density function of x and y is*

$$f(x, y) = 2xe^{-y}, 0 < x < 1, y > 0$$
$$= 0 \text{ elsewhere,}$$

Find the distribution of x + y.



Sol. Put $u = x + y, v = y$

$\therefore \quad x = u - v, y = v$

$y > 0 \Rightarrow v > 0$

$0 < x < 1 \Rightarrow 0 < u - v < 1$

$\therefore$ u and v vary in shaded portion.

Now
$$\frac{1}{J} = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$$

$\therefore$
$$J = 1$$

$\therefore$ The joint density function of u and v is given by

$$F(u, v) = 2(u-v)e^{-v}$$

To find the density function of u divide the shaded portion in two parts marked I and II by the line $x = 1$.

In part I, $0 < u \leq 1$ and for given u, v varies from 0 to u and in part II, $u \geq 1$ and for given u, v varies from $u-1$ to u.

$\therefore$ Density function of u is given by

$$f_u(u) = \int_0^u 2(u-v)e^{-v}\,dv$$

$$= 2\left[-(u-v)e^{-v} + e^{-v}\right]_{v=0}^{u}$$

$$= 2(e^{-u} + u - 1) \text{ for } 0 < u \leq 1$$

and for $u \geq 1$,

$$g(u) = 2\int_{u-1}^{u} (u-v)e^{-v}\,dv$$

$$= 2\left[-(u-v)e^{-v} + e^{-v}\right]_{u-1}^{u}$$

$$= 2\{e^{-u} + e^{-(u-1)}) - e^{-(u-1)}\}$$

$$= 2e^{-u}$$

Ex. 11-33. *If x, y are independent standard normal variates, show that $\dfrac{x}{y}$ follow Cauchy distribution.*

Sol. Joint dist. of x and y is

$$dP = \left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx\right\} \cdot \left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\,dy\right\}$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy.$$

Put $\frac{x}{y}$

Then both u, v vary from $-\infty$ t

$x = uv, \quad y = v.$

$\therefore$

$\therefore$ Joint dist. of u, v is

dP

$\therefore$ Dist. of u is given by

dP

Ex. 11-34. *If (x, y) have a biv*

variates with correlation co-efficien

Sol. Joint dist. of (x, y) is

dp

Put u

$\therefore$ x

y

portion in two parts marked I and II

0 to u and in part II, $u \geq 1$ and for

$e^{-v}\Big]_{v=0}^{u}$

or $0 < u \leq 1$

$e^{-v}\Big]_{u-1}^{u}$

$- e^{-(u-1)}\}$

iriates, show that $\frac{x}{y}$ follow Cauchy

$\left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy\right\}$

ly.

Put $\qquad \frac{x}{y} = u, \quad y = v$

Then both u, v vary from $-\infty$ to $+\infty$.

$x = uv, \qquad y = v.$

$$\therefore \qquad J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

$\therefore$ Joint dist. of u, v is

$$dP = \frac{1}{2\pi} e^{-\frac{1}{2}v^2(1+u^2)} |v| \, du \, dv$$

$\therefore$ Dist. of u is given by

$$dP = \frac{du}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2(1+u^2)} |v| \, dv$$

$$\frac{du}{\pi} \int_{0}^{\infty} e^{-\frac{1}{2}v^2(1+u^2)} v \, dv$$

$$= \frac{du}{\pi} \left\{ \frac{e^{-\frac{1}{2}v^2(1+u^2)}}{-(1+u^2)} \right\}_0^{\infty}$$

$$= \frac{1}{\pi} \frac{du}{1+u^2}.$$

Ex. 11-34. *If (x, y) have a bivariate normal distribution s.t. x, y are standard normal variates with correlation co-efficient ρ between them, show that x/y follow Cauchy dist.*

Sol. Joint dist. of (x, y) is

$$dp = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{x^2 - 2\rho xy + y^2\}} dx \, dy$$

Put $\qquad u = \frac{x}{y}, \quad v = y$

$\therefore \qquad x = uv, \quad y = v$

$$\therefore \qquad J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\[2ex] \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v.$$

$\therefore$ Joint dist. of u, v is

$$dP = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{v^2}{2(1-\rho^2)}\{u^2-2\rho u+1\}} |v|\, du\, dv$$

$\therefore$ Dist. of u is

$$dP = \frac{du}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2(1-\rho^2)}\{u^2-2\rho u+1\}} |v|\, dv$$

$$= \frac{du}{\pi\sqrt{1-\rho^2}} \int_{0}^{\infty} e^{-\frac{v^2}{2(1-\rho^2)}\{u^2-2\rho u+1\}} v dv$$

Put $$\frac{v^2}{2(1-\rho^2)}(u^2-2\rho u+1) = z$$

$$\Rightarrow \qquad v\, dv = \frac{(1-\rho^2)\, dz}{u^2-2\rho\, u+1}$$

$$\therefore \qquad dP = \frac{\sqrt{1-\rho^2}}{\pi(u^2-2\rho u+1)} \int_0^{\infty} e^{-z}\, dz$$

$$= \frac{\sqrt{1-\rho^2}}{\pi\{(u-\rho)^2+(1-\rho^2)\}}$$

which is Cauchy distribution.

EXERCISES

1. Calculate the missing entries in the following bivariate distribution table :

$y \rightarrow$ / $x \downarrow$	0	1	2	Total
1	0	—	—	·2
2	—	·3	—	·4
3	0	—	·1	—
Total	·1	—	·2	

2. The joint density function of a b

$f(x, y) =$

(i) Find marginal and conditio

(ii) Are x and y independent ?

3. Obtain the marginal and conditio

$f(x, y) = c$ (

$= 0$

4. If the joint density function of a

$f(x, y) = c$

$= 0$

Find (i) $P(x > y)$

(ii) $P(x + y < 1)$

5. Given the following bivariate pr

$x \rightarrow$

$y \downarrow$

0

1

2

Find (i) marginal distributions

(ii) the conditional distribu

6. Random variables x, y have the

$f(x, y) = \dfrac{}{2\pi\sqrt{}}$

Find the marginal densities of x,

7. Two stochastic variables x and y as follows

$y \rightarrow$

$x \downarrow$

1

2

3

Find $E(x)$, $E(xy)$ and $E(x + y)$.

8. Given the following freq distrib the correlation co-efficient :

$\frac{v^2}{1-\rho^2)}\{u^2-2\rho u+1\}$ $|v|\,du\,dv$

$\frac{v^2}{2(1-\rho^2)}\{u^2-2\rho u+1\}$ $|v|\,dv$

$\frac{v^3}{(1-\rho^2)}\{u^2-2\rho u+1\}$ $v\,dv$

$2\rho u+1)=z$

$\int_0^\infty e^{-z}\,dz$

$-\rho^2)\}$

ıriate distribution table :

Total

·2

·4

2. The joint density function of a bivariate distribution is given by

$$f(x,y)=\begin{cases} c\,xy, & 0<y<x<1 \\ 0 & \text{otherwise}\end{cases}$$

(*i*) Find marginal and conditional distributions.

(*ii*) Are x and y independent ? Why ?

3. Obtain the marginal and conditional probability functions for the following distribution

$$\begin{aligned} f(x,y) &= c\,(2-x-y), && 0<y<x<1 \\ &= 0 && \text{elsewhere.}\end{aligned}$$

4. If the joint density function of a bivariate distribution is given by

$$\begin{aligned} f(x,y) &= c.e^{-(x+y)}, && x\ge 0,\ y\ge 0 \\ &= 0 && \text{elsewhere}\end{aligned}$$

Find (*i*) $P(x>y)$

(*ii*) $P(x+y<1)$

5. Given the following bivariate probability distribution.

$y\downarrow$ \ $x\rightarrow$	-1	0	1
0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$
1	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
2	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{1}{15}$

Find (*i*) marginal distributions of x and y

(*ii*) the conditional distribution of x given $y=1$.

6. Random variables x, y have the joint probability density

$$f(x,y)=\frac{1}{2\pi\sqrt{1-k^2}}\exp\left\{-\frac{x^2-2kxy+y^2}{2(1-k^2)}\right\}$$
$$-\infty<x,\ y<\infty.$$

Find the marginal densities of x, y and also their co-efficient of correlation.

7. Two stochastic variables x and y take the values 1, 2, 3 only and their probabilities are as follows

$x\downarrow$ \ $y\rightarrow$	1	2	3
1	0·1	0·1	0·1
2	0·1	0·2	0·1
3	0·1	0·1	0·1

Find $E(x)$, $E(xy)$ and $E(x+y)$.

8. Given the following freq distribution, find the mean values, variances, covariance and the correlation co-efficient :

$x \rightarrow$ / $y \downarrow$	-1	1	Total
-1	20	80	100
1	80	20	100
Total	100	100	200

State, with reasons but without making any calculations, what shall be the corresponding results if the two frequencies in each row are interchanged.

9. Let the joint density f^n of x and y be given by

$$f(x,y)=\begin{cases}8xy & , \quad 0<x<y<1\\ 0 & , \quad \text{elsewhere.}\end{cases}$$

Find $E(y/x)$, $E(xy/x)$, $\text{var}(y/x)$.

$$\left\{\textbf{Ans:}\ \frac{2}{3}\,\frac{1+x+x^2}{1+x},\ \frac{2}{3}\,\frac{x(1+x+x^2)}{1+x},\ \frac{1+2x-6x^2+2x^3+x^4}{18(1+x)^2}\right.$$

10. For the bivariate distribution

$$f(x,y)=\frac{1}{8}(6-x-y) \qquad 0<x<2;\ 2<y<4.$$

Find $E(y/x)$, $\quad E(y^2/x)$, $\quad \text{var}\,(y/x)$, $E(xy/x)$.
Also show that

$$E(y)=E[E(y/x)].$$

11. The joint *p.d.f.* of x and y is given by

$$f(x,y)=\begin{cases}2 & , \quad 0<x<y<1\\ 0 & , \quad \text{otherwise.}\end{cases}$$

Show that the conditional variance of x given y is $\dfrac{y^2}{12}$.

12. x_1, x_2 are independent random variables having the same Cauchy distribution

$$f(x)=\frac{1}{\pi(1+x^2)},\ -\infty<x<\infty.$$

Show that $y=x_1+x_2$ has a Cauchy distribution.

13. Let x and y be independent variates which are uniformly distributed over the unit interval (0, 1), find the distribution f^n and p.d.f of $z=x+y$. Is z a uniformly distributed variable ?

14. If z is $N(0, 1)$, show that the p.d.f (*i.e.*, prob. density f^n) of $|z|$ is

$$f(z)=\left(\frac{2}{\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}z^2},\ x\geq 0.$$

Now let x and y be independent $N(0, 1)$ variates. Show that $u=\dfrac{x}{|y|}$ has the distribution specified by

$g(u) = \ldots$

15. Suppose that x_1 and x_2 are inc
normal distribution, then :
(*i*) Find the joint distribution of

x_1

(*ii*) Show that $2x_1x_2$ and $x_2{}^2$ -

16. x and y are independent random
has the chi-square distribution w

Find the prob. density of $\dfrac{x}{\sqrt{y/n}}$

Hence, deduce that if x, z are in

then $\dfrac{x}{z}$ has a Cauchy distributic

17. Let

$f(x)$

be *p.d.f.* of the random variable
density f^n of $y=x^2$.

18. If x and y are two independent ra

and $g($

Find the probability distribution

19. If the joint density function of t
density function of $u=x+y$ is

20. If x and y are independent Bi
respectively,

Total

100
100
200

ions, what shall be the corresponding
changed.

$x < y < 1$
where.

$5x^2 + 2x^3 + x^4$
$(1+x)^2$

$2; 2 < y < 4$.

$y < 1$
ise.

$\frac{y^2}{12}$.

e same Cauchy distribution

$< \infty$.

ly distributed over the unit interval
$+ y$. Is z a uniformly distributed

f^n) of $|z|$ is

).

w that $u = \frac{x}{|y|}$ has the distribution

$$g(u) = \frac{1}{\pi(1+u^2)}, -\infty < u < \infty.$$

15. Suppose that x_1 and x_2 are independent random variables each having a standard normal distribution, then :

(*i*) Find the joint distribution of

$$\frac{x_1 + x_2}{\sqrt{2}} \text{ and } \frac{x_2 - x_1}{\sqrt{2}}$$

(*ii*) Show that $2x_1x_2$ and $x_2{}^2 - x_1{}^2$ have the same distribution.

16. x and y are independent random variables. x has the standard normal distribution and y has the chi-square distribution with n degrees of freedom

Find the prob. density of $\dfrac{x}{\sqrt{y/n}}$

Hence, deduce that if x, z are independent and have the standard normal distribution, then $\dfrac{x}{z}$ has a Cauchy distribution.

17. Let

$$f(x) = \begin{cases} \frac{1}{2}, & -1 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

be *p.d.f.* of the random variable x. Find the distribution function and the probability density f^n of $y = x^2$.

18. If x and y are two independent random variables with density functions

$$f(x) = e^{-x}, \quad x \geq 0$$

and

$$g(y) = 2e^{-2y}, \quad y \geq 0.$$

Find the probability distribution of $u = \dfrac{x}{y}$.

$$\left\{ \textbf{Ans. } \frac{1}{(u+2)^2} \right\}$$

19. If the joint density function of two random variates x and y is $f(x, y)$, show that the density function of $u = x + y$ is

$$\int_{-\infty}^{\infty} f(v, u-v)\,dv$$

20. If x and y are independent Binomial variates with parameters $3; \frac{1}{2}$ and $2; \frac{1}{2}$ respectively,

find $P(x = y)$

$\left(\textbf{Ans. } 10/2^5\right)$

21. If X and Y are two random variables and $E\{Y|X = x\} = \mu$, where μ does not depend upon X, show that

$$\text{var}(Y) = E\{\text{var}(Y/X)\}$$

22. Define moment generating function of $Y/X = x$. Does $M_y(t) = E\{M_{Y/X}(t)\}$.

23. If x and y be two independent random variates, does $E\{Y/X = x\}$ depend upon x.

24. If the joint moment generating function of x, y is given by

$$M_{x,y}(t_1, t_2) = \exp\left\{\left(t_1^2 + t_2^2\right)/2\right\},$$

find the distribution of x.

25. Let X, Y be random variables with joint probability density function $f_{X,Y}(x, y)$. Let $U(X)$ and $V(Y)$ be functions of X, Y respectively. Then show that

$$E\{U(X)V(Y)|X = x\} = U(x)E\{V(Y)|X = x\}.$$

26. The trinomial distribution (multinomial with $k+1=3$) of two random variables X and Y is given by

$$f_{X,Y}(x, y) = \frac{n!}{x!\,y!\,(n-x-y)!}\,p^x q^y (1-p-q)^{n-x-y}$$

for $x, y = 0, 1, \ldots\ldots n$ and $x + y \le n$ where $0 \le p, 0 \le q$ and $0 \le p + q \le 1$. Find:

(*i*) marginal distribution of Y

(*ii*) the conditional distribution of X given Y and obtain its expected value.

(*iii*) $\rho_{x,y}$.

(See section : 10·8)

□□□

(**Ans.** $10/2^5$)

$= x\} = \mu$, where μ does not depend

$X)\}$

. Does $M_y(t) = E\{M_{Y/X}(t)\}$.

loes $E\{Y/X = x\}$ depend upon x.

is given by

$t_2^2)/2\}$,

lity density function $f_{X,Y}(x,y)$. Let

ely. Then show that

$V(Y)|X = x\}$.

$+1 = 3)$ of two random variables X

$(1-p-q)^{n-x-y}$

$, 0 \le q$ and $0 \le p+q \le 1$. Find:

l obtain its expected value.

(See section : 10·8)

□□□

12

Special Continuous Distributions

12.1. Uniform Distribution

It is a continuous dist. given by

$$dF = \frac{1}{b-a}\,dx, \quad a \le x \le b$$

a, b are called the parameters of the dist. We have $a < b$.

Distribution f^n is

$$F(x) = \frac{1}{b-a}\int_a^x dx$$

$$= \frac{x-a}{b-a}.$$

12.1.1. Mean and Variance

$$\bar{x} = E(x)$$

$$= \frac{1}{b-a}\int_a^b x\,dx$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

$$\mu'_2(0) = \frac{1}{b-a}\int_a^b x^2\,dx$$

$$= \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{b^2+a^2+ab}{3}$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \bar{x}^2$$

$$= \frac{b^2+a^2+ab}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{(b-a)^2}{12}$$

12.1.2. M.G.F.

$$M_0(t) = E(e^{tx})$$

$$= \frac{1}{b-a}\int_a^b e^{tx}\,dx$$

$$= \frac{e^{tb}-e^{ta}}{t(b-a)}.$$

12.2. Gamma Distribution

Let x be a continuous random variate with probability density function defined as below :

$$f(x) = \frac{1}{\Gamma(\lambda)} e^{-x} x^{\lambda-1},\ \lambda > 0,\ 0 < x < \infty$$

x is called **Gamma variate** with parameter λ and is referred to as a $\gamma(\lambda)$ variate. The distribution of x is called a **Gamma distribution.**

12.2.1. Moments about mean and β, γ co-efficients

$$\mu'_r(0) = E(x^r)$$

$$= \frac{1}{\Gamma(\lambda)}\int_0^\infty x^r e^{-x} x^{\lambda-1}\,dx$$

$$= \frac{1}{\Gamma(\lambda)}\int_0^\infty e^{-x} x^{(\lambda+r)-1}\,dx$$

$$= \frac{1}{\Gamma(\lambda)}\Gamma(\lambda+r)$$

$$= (\lambda+r-1)(\lambda+r-2)\ldots(\lambda)$$

$$\text{mean} = \mu'_1(0) = \lambda$$

$$\mu'_2(0) = \lambda(\lambda+1)$$

$$\mu'_3(0) = \lambda(\lambda+1)(\lambda+2)$$

$$\mu'_4(0) = \lambda(\lambda+1)(\lambda+2)(\lambda+3)$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \{\mu'_1(0)\}^2$$

$$= \lambda(\lambda+1) - \lambda^2$$

$$= \lambda$$

$$\mu_3 = \mu'_3(0) - 3\mu'_2(0)\mu'_1(0) + 2\{\mu'_1(0)\}^3$$

$$= \lambda(\lambda+1)(\lambda+2) - 3\lambda^2(\lambda+1) + 2\lambda^3$$

$$= \lambda^3 + 3\lambda^2 + 2\lambda - 3\lambda^3 - 3\lambda^2 + 2\lambda^3$$

$$= 2\lambda$$

$$\mu_4 = \mu'_4(0) - 4\mu'_3(0)\mu'_1(0) + 6\mu'_2(0)\{\mu'_1(0)\}^2 - 3\{\mu'_1(0)\}^4$$

$$= \lambda(\lambda+1)(\lambda+2)(\lambda+3) - 4\lambda^2(\lambda+1)(\lambda+2) + 6\lambda^3(\lambda+1) - 3\lambda^4$$

$$= \lambda\{\lambda^3 + 6\lambda^2 + 11\lambda + 6) - 4\lambda^2(\lambda^2+3\lambda+2) + 6\lambda^4 + 6\lambda^3 - 3\lambda^4$$

$$= 3\lambda^2 + 6\lambda$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4\lambda^2}{\lambda^3} = \frac{4}{\lambda}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{6}{\lambda}$$

$$\gamma_1 = \sqrt{\beta_1} = \ldots$$

$$\gamma_2 = \beta_2 - 3 = \ldots$$

12.2.2. Mode

Density function is

$$f(x) = \ldots$$

$$\therefore \quad f'(x) = \ldots$$

$$\therefore \quad f'(x) = 0$$

for $x = 0, f(x) = 0$ which is mini…

$$\therefore \quad \text{for } x = \lambda \ldots$$

$$\therefore \quad \text{Mode} = \lambda \ldots$$

12.2.3. M.G.F. of Gamma Distribu…

$$M_0(t) = E\{\ldots$$

$$M_{\bar{x}}(t^1) = E\{\ldots$$

$$= e^{-\lambda}\ldots$$

$$= e^{-\lambda}\ldots$$

12.2.4. Cumulative Generating Fu…

$$K_0(t) = \log\ldots$$

$$= -\lambda\ldots$$

$$= \lambda\ldots$$

$$\therefore \quad k_1(0) = \lambda, \ldots$$

In general,

$$\frac{k_r}{r!} = \text{co}\ldots$$

$$= \frac{\lambda}{r}$$

$$\Rightarrow \quad k_r = \lambda(\ldots$$

y density function defined as below :

$0, 0 < x < \infty$

referred to as a $\gamma(\lambda)$ variate. The

)

$\lambda + 3)$

$2\{\mu'_1(0)\}^3$

$1) + 2\lambda^3$

$- 2\lambda^3$

$\mu'_2(0)\{\mu'_1(0)\}^2 - 3\{\mu'_1(0)\}^4$

${}^2(\lambda+1)(\lambda+2) + 6\lambda^3(\lambda+1) - 3\lambda^4$

${}^2(\lambda^2+3\lambda+2) + 6\lambda^4 + 6\lambda^3 - 3\lambda^4$

$$\gamma_1 = \sqrt{\beta_1} = \frac{2}{\sqrt{\lambda}}$$

$$\gamma_2 = \beta_2 - 3 = \frac{6}{\lambda}.$$

12.2.2. Mode

Density function is

$$f(x) = \frac{1}{\Gamma(\lambda)} e^{-x} . x^{\lambda-1}$$

$$\therefore \quad f'(x) = \frac{1}{\Gamma(\lambda)} e^{-x} . x^{\lambda-2} \{(\lambda-1)-x\}$$

$$\therefore \quad f'(x) = 0 \Rightarrow x = \lambda - 1, 0$$

for $x = 0, f(x) = 0$ which is minimum value of $f(x)$.

$$\therefore \quad \text{for } x = \lambda - 1, f(x) \text{ is maximum}$$

$$\therefore \quad \text{Mode} = \lambda - 1.$$

12.2.3. M.G.F. of Gamma Distribution

$$M_0(t) = E\{e^{tx}\}.$$

$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{tx} e^{-x} x^{\lambda-1} dx$$

$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{-x(1-t)} x^{\lambda-1} dx$$

$$= \frac{1}{\Gamma(\lambda)} \cdot \frac{1}{(1-t)^\lambda} \int_0^\infty e^{-y} y^{\lambda-1} dy,$$

where $y = (1-t)x, |t| < 1$

$$= \frac{1}{\Gamma(\lambda)} \cdot \frac{1}{(1-t)^\lambda} \Gamma(\gamma) = \frac{1}{(1-t)^\lambda}$$

$$M_{\bar{x}}(t^l) = E\{e^{(x-\lambda)t}\}$$

$$= e^{-\lambda t} M_0(t)$$

$$= e^{-\lambda t}(1-t)^{-\lambda}, \qquad |t| < 1.$$

12.2.4. Cumulative Generating Function and Cumulants

$$K_0(t) = \log M_0(t)$$

$$= -\lambda \log(1-t), \qquad |t| < 1$$

$$= \lambda\left\{t + \frac{t^2}{2} + \frac{t^3}{3} + \ldots..\right\}$$

$$\therefore \quad k_1(0) = \lambda, k_2 = \lambda \text{ etc.}$$

In general,

$$\frac{k_r}{r!} = \text{co-eff. of } t^r$$

$$= \frac{\lambda}{r}$$

$$\Rightarrow \quad k_r = \lambda(r-1)!$$

12.2.5. Additive Property of Gamma Variates

Theorem. The sum of any finite number of independent Gamma variates is a Gamma variate.

Proof. Let $x_1\ x_2\ \ldots.\ x_n$ be independent Gamma variates with parameters $\lambda_1, \lambda_2 \ldots \lambda_n$ respectively.

Then $\quad M_0(t) \text{ of } x_i = (1-t)^{-\lambda_i} \qquad (i = 1, 2 \ldots n)$

Let $\quad X = x_1 + x_2 \ldots\ldots + x_n$

Then
$$M_0(t) \text{ of } X = E\{e^{t(x_1 + \ldots + x_n)}\}$$
$$= E\{e^{tx_1}\}\,.\,E\{e^{tx_2}\}\ldots.E\{e^{tx_n}\}$$
$$= (1-t)^{-\lambda_1}\,.\,(1-t)^{-\lambda_2}\ldots.(1-t)^{-\lambda_n}$$
$$= (1-t)^{-(\lambda_1 + \ldots + \lambda_n)}$$

which is the m.g.f. of a $\gamma(\lambda_1 + \ldots. + \lambda_n)$

$\therefore \quad X$ is a $\gamma(\lambda_1 + \ldots. + \lambda_n)$.

12.2.6. Limiting Form of Gamma Distribution

Theorem. Show that as $\lambda \to \infty$, Gamma Distribution tends to normal distribution.

Proof. Let x be a $\gamma(\lambda)$. Then $\bar{x} = \lambda$, var $(x) = \lambda$,

Let
$$z = \frac{x-\lambda}{\sqrt{\lambda}}$$

$$\therefore \quad M_0(t) \text{ of } z = E\left\{e^{t\left(\frac{x-\lambda}{\sqrt{\lambda}}\right)}\right\}$$
$$= e^{-\sqrt{\lambda}.t}\ E\left\{e^{\left(\frac{t}{\sqrt{\lambda}}\right)x}\right\}$$
$$= e^{-\sqrt{\lambda}.t}\left(1 - \frac{t}{\sqrt{\lambda}}\right)^{-\lambda}$$

$$\therefore \quad \log\{M_0(t) \text{ of } z\} = -\sqrt{\lambda}\,t - \lambda \log\left(1 - \frac{t}{\sqrt{\lambda}}\right)$$
$$= -\sqrt{\lambda}\,t + \lambda\left\{\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{3\lambda\sqrt{\lambda}} + \ldots..\right\}$$
$$= \frac{1}{2}t^2 + \text{terms containing } \lambda \text{ in the denominator}$$
$$\to \frac{1}{2}t^2 \text{ as } \lambda \to \infty$$

$$\therefore \quad M_0(t) \text{ of } z \to e^{\frac{1}{2}t^2} \quad as\ \lambda \to \infty$$

which is the m.g.f. of a standard normal variate.

$\therefore$ In the limiting case, z and hence x is a normal variate.

12.3. Beta Distribution of First Ki

Let x be a continuous random vari

$f(x) = $ β(

x is known as a Beta variate of firs $\beta_1(m, n)$ variate.

The distribution of x is called Be

12.3.1. Moments and Harmonic M

$\therefore \quad \mu'_r(0) = E($

Mean $= \mu'$

$\mu'_2(0) =$

$\mu_2 = \mu'$

The harmonic mean H is given

$\frac{1}{H} = E$

pendent Gamma variates is a Gamma

variates with parameters $\lambda_1, \lambda_2 \dots \lambda_n$

$(i = 1, 2 \dots n)$

$\dots E\{e^{tx_n}\}$

$\lambda_2 \dots (1-t)^{-\lambda_n}$

bution tends to normal distribution.

$1 - \frac{t}{\sqrt{\lambda}}\Big)$

$+ \frac{t^2}{2\lambda} + \frac{t^3}{3\lambda\sqrt{\lambda}} + \dots \Big\}$

ntaining λ in the denominator

nal variate.

12.3. Beta Distribution of First Kind

Let x be a continuous random variate with probability density function defined as below :

$$f(x) = \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}, \; m, n > 0, \; 0 < x < 1$$

x is known as a Beta variate of first kind with parameters m and n. *It is referred to as* $\beta_1(m, n)$ variate.

The distribution of x is called Beta distribution of first kind.

12.3.1. Moments and Harmonic Mean

$$\therefore \quad \mu'_r(0) = E(x^r)$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty x^r . x^{m-1}(1-x)^{n-1}\, dx$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty x^{m+r-1}(1-x)^{n-1}\, dx$$

$$= \frac{\beta(m+r,n)}{\beta(m,n)}$$

$$= \frac{\Gamma(m+r)\Gamma(m+n)}{\Gamma(m+n+r)\Gamma(m)}$$

$$= \frac{(m+r-1)(m+r-2)\dots(m)}{(m+n+r-1)\dots(m+n)}$$

$$\therefore \quad \text{Mean} = \mu'_1(0) = \frac{m}{m+n}$$

$$\mu'_2(0) = \frac{m(m+1)}{(m+n)(m+n+1)}$$

$$\mu_2 = \mu'_2(0) - \{\mu'_1(0)\}^2$$

$$= \frac{m(m+1)}{(m+n)(m+n+1)} - \left\{\frac{m}{m+n}\right\}^2$$

$$= \frac{mn}{(m+n)^2(m+n+1)}$$

The harmonic mean H is given by

$$\frac{1}{H} = E\left(\frac{1}{x}\right)$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty \frac{1}{x} \cdot x^{m-1}(1-x)^{n-1}\, dx$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty \cdot x^{m-2}(1-x)^{n-1}\, dx$$

$$= \frac{\beta(m-1,n)}{\beta(m,n)}$$

$$= \frac{\Gamma(m-1)\Gamma(m+n)}{\Gamma(m)\Gamma(m+n-1)}$$

$$= \frac{m+n-1}{m-1}$$

$$\therefore \qquad H = \frac{m-1}{m+n-1}.$$

12.3.2. Mode of Beta-distribution of first kind

We have

$$f(x) = \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}, \qquad 0<x<1 \qquad m,n>0$$

If $m<1$:

$f(x) \to \infty$ as $x \to 0^+$

$\therefore$ $x = 0$ is the modal value.

If $n<1$:

$f(x) \to \infty$ as $x \to 1^-$

$\therefore$ $x = 1$ is the mode.

If $m<1$; $n<1$:

both $x = 0, 1$ are modes

If $m=1$, $n=1$:

$$f(x) = \frac{1}{\beta(m,n)} = 1$$

$\therefore$ Each value of x is modal value.

If $m>1$, $n=1$:

$$f(x) = \frac{x^{m-1}}{\beta(m,1)}$$

Here $f(x)$ increases as x increases.

$\therefore$ f is maximum at $x = 1$

$\therefore$ $x = 1$ is the mode

If $m=1$, $n>1$:

$$f(x) = \frac{(1-x)^{n-1}}{\beta(1,n)}$$

$f(x)$ decreases as $x \to 1^-$

$\therefore$ f is maximum at $x = 0$

$\therefore$ $x = 0$ is the mode.

If $m>1$, $n>1$:

$$f(x) = \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}$$

$$f'(x) = \frac{1}{\beta(m,n)}\left\{(m-1)x^{m-2}(1-x)^{n-1} - (n-1)x^{m-1}(1-x)^{n-2}\right\}$$

$$= \frac{1}{\beta(m,n)} x^{m-2}(1-x)^{n-2}\{(m-1)(1-x)-(n-1)x\}$$

Mode is given by $f'(x) = 0$

$\Rightarrow$ $(m-1)(1-x)-(n-1)x =$

$x = \frac{m-}{m+n}$

(Obviously this value lies in (0, 1

Ex. 12-1. *If x has a* $\beta_1(m, n)$ *di*

Sol. Let x be a Beta variate of fi

Then $E(1/x) = \frac{m-}{m+n}$

$\Rightarrow$ $E(1/x) = 1$

which is not as $n>0$

$\therefore$ $E(1/x) \neq 1$

12.4. Beta Distribution of Second I

Let x be a continuous random vari

$f(x) = \frac{}{\beta(}$

x is known as Beta variate of secon

$\beta_2(m, n)$ *variate.* The distribution of

Remarks : If x is a $\beta_2(m, n)$ va

$y = \frac{1}{1+}$

is a $\beta_1(m, n)$ variate.

12.4.1. Moments and Harmonic M

$\mu'_r(0) = E($

$\therefore$ Mean $= \mu'_1(0) =$

$\mu'_2(0) =$

$)^{n-1}, \quad 0 < x < 1 \qquad m, n > 0$

$x)^{n-1}$

$^{-2}(1-x)^{n-1} - (n-1)x^{m-1}(1-x)^{n-2}\}$

$)^{n-2}\{(m-1)(1-x)-(n-1)x\}$

$$\Rightarrow \quad (m-1)(1-x)-(n-1)x = 0$$

$$x = \frac{m-1}{m+n-2}$$

(Obviously this value lies in (0, 1))

Ex. 12-1. *If x has a* $\beta_1(m, n)$ *distribution, can* $E(1/x)$ be unity ?

Sol. Let x be a Beta variate of first kind with parameters m and n.

Then $$E(1/x) = \frac{m-1}{m+n-1}$$

$$\Rightarrow \quad E(1/x) = 1 \text{ iff } n = 0$$

which is not as $n > 0$

$$\therefore \quad E(1/x) \neq 1$$

12.4. Beta Distribution of Second Kind

Let x be a continuous random variate with probability density function defined as below :

$$f(x) = \frac{1}{\beta(m,n)} \frac{x^{m-1}}{(1+x)^{m+n}}, \; m, n > 0, \; 0 < x < \infty$$

x is known as Beta variate of second kind with parameters, m and n. *It is referred to as* $\beta_2(m, n)$ *variate*. The distribution of x is called Beta distribution of second kind.

Remarks : If x is a $\beta_2(m, n)$ variate, then

$$y = \frac{1}{1+x}$$

is a $\beta_1(m, n)$ variate.

12.4.1. Moments and Harmonic Mean

$$\mu'_r(0) = E(x^r)$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty x^r \cdot \frac{x^{m-1}}{(1+x)^{m+n}} \, dx$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty \frac{x^{m+r-1}}{(1+x)^{m+n}} \, dx$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty \frac{x^{m+r-1}}{(1+x)^{(m+r)+(n-r)}} \, dx$$

$$= \frac{1}{\beta(m,n)} \beta(m+r, n-r)$$

$$= \frac{\Gamma(m+r)\Gamma(n-r)}{\Gamma(m)\Gamma(n)}$$

$$= \frac{(m+r-1)(m+r-2)....(m)}{(n-1)(n-2)....(n-r)}$$

$$\therefore \quad \text{Mean} = \mu'_1(0) = \frac{m}{n-1}$$

$$\mu'_2(0) = \frac{m(m+1)}{(n-1)(n-2)}$$

$$\mu_2 = \mu'_2(0) - \{\mu'_1(0)\}^2$$

$$= \frac{m(m+1)}{(n-1)(n-2)} - \left(\frac{m}{n-1}\right)^2$$

$$= \frac{m(m+n-1)}{(n-1)^2(n-2)}$$

The harmonic mean H is given by

$$\frac{1}{H} = E\left(\frac{1}{x}\right)$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty \frac{1}{x} \frac{x^{m-1}}{x(1+x)^{m+n}}\, dx$$

$$= \frac{1}{\beta(m,n)} \int_0^\infty \frac{x^{(m-1)-1}}{(1+x)^{(m-1)+(n+1)}}\, dx$$

$$= \frac{1}{\beta(m,n)}\, \beta(m-1, n+1)$$

$$= \frac{\Gamma(m-1)\Gamma(n+1)}{\Gamma(m)\Gamma(n)}$$

$$= \frac{n}{m-1}$$

$$H = \frac{m-1}{n}.$$

12.4.2. Mode of Beta dist. of second kind

$$f(x) = \frac{1}{\beta(m,n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}},\ m, n > 0,\ 0 < x < \infty$$

$$f'(x) = \frac{1}{\beta(m,n)} \cdot \left\{ \frac{(m-1)x^{m-2}(1+x)^{m+n} - (m+n)x^{m-1}(1+x)^{m+n-1}}{(1+x)^{2m+2n}} \right\}$$

Put $f'(x) = 0$

$$\Rightarrow \quad (m-1)\, x^{m-2}(1+x)^{m+n} - (m+n)\, x^{m-1}(1+x)^{m+n-1} = 0$$

$$\Rightarrow \quad x^{m-2}(1+x)^{m+n-1}\{(m-1)(1+x) - (m+n)\,x\} = 0$$

Neglecting $x = 0$,

$$(m-1)(1+x) - (m+n)\,x = 0$$

$$\Rightarrow \quad x = \frac{m-1}{n+1}.$$

This exists only when $m > 1$.

12.5. Exponential Distribution

It is a continuous distribution given by

$$dF = \alpha e^{-\alpha x}, x > 0,$$

where α is parameter and $\alpha > 0$, x is called exponential variate with parameter α, we have

$$\mu'_r(0) = E(x^r)$$

12.5.2. If $x_1, x_2, ... x_n$ be indepen
respectively, then

has exponential distribution with p

Proof. We have

$$= \alpha \int_0^\infty x^r e^{-\alpha x}\, dx$$

Put $\quad \alpha x = t$

$$\Rightarrow \quad dx = \frac{dt}{\alpha}$$

$$= \frac{1}{\alpha^r} \int_0^\infty t^r e^{-t}\, dt$$

$$= \frac{r!}{\alpha^r}.$$

$$\therefore \quad \bar{x} = \mu_1'(0) = \frac{1}{\alpha}$$

$$\mu_2'(0) = \frac{2}{\alpha^2}$$

$$\therefore \quad \text{var}(x) = \mu_2'(0) - \bar{x}^2$$

$$= \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

$$M_0(t) = E(e^{tx})$$

$$= \alpha \int_0^\infty e^{tx} e^{-\alpha x}\, dx$$

$$= \alpha \int_0^\infty e^{(t-\alpha)x}\, dx$$

$$= \frac{\alpha}{t-\alpha} \left| e^{(t-\alpha)x} \right|_0^\infty$$

$$= \frac{\alpha}{\alpha - t}, \text{ if } t < \alpha.$$

12.5.2. If $x_1, x_2, \ldots\ldots x_n$ be independent exponential variates with parameters $\alpha_1, \alpha_2, \ldots\ldots \alpha_n$ respectively, then

$$x = \min(x_1, x_2, \ldots\ldots x_n)$$

has exponential distribution with parameter $\sum_{i=1}^{n} \alpha_i$.

Proof. We have

$$P(x \le X) = 1 - P(x > X)$$

$$= 1 - P\{\min(x_1, x_2, \ldots x_n) > X\}$$

$$= 1 - P\{(x_i > X)\}/\forall i\}$$

$$= 1 - P(x_1 > X) \,.\, P(x_2 > X) \ldots\ldots P(x_n > X)$$

$$(\because x_i\text{'s are independent}) \qquad \ldots(1)$$

Now $\quad P(x_i > X) = 1 - P(x_i \le X)$

$$= 1 - \int_0^X \alpha_i e^{-\alpha_i x_i}\, dx_i$$

$$= e^{-\alpha_i X}$$

$\therefore \quad (1) \Rightarrow \quad P(x \le X) = 1 - e^{-\alpha_1 X} \cdot e^{-\alpha_2 X} e \ldots\ldots\ldots^{-\alpha_n X}$

$$= 1 - e^{-(\Sigma\alpha_i)X}$$

$\therefore$ Cumulative distribution f^n of x is

$$F(x) = 1 - e^{-(\Sigma\alpha_i)x}$$

$\therefore$ Density f^n of x is given by

$$f(x) = F'(x) = (\Sigma\alpha_i)\, e^{-(\Sigma\alpha_i)x}$$

$\therefore$ x has exponential distribution with parameter

$$\sum_{i=1}^{n} \alpha_i .$$

12.6. Double Exponential or Laplace's Distribution

It is a continuous distribution given by

$$df = \frac{1}{2\beta} \exp\left\{-\frac{|x-\alpha|}{\beta}\right\}, -\infty < x < \infty$$

where $\beta > 0$

α, β are called parameters of the distribution.

We have

$$\mu'_r(0) = E(x^r)$$

$$= \frac{1}{2\beta} \int_{-\infty}^{\infty} x^r \cdot e^{-\frac{|x-\alpha|}{\beta}}\, dx$$

Put $\dfrac{x-\alpha}{\beta} = y$

$$= \frac{1}{2} \int_{-\infty}^{\infty} (\alpha+\beta y)^r e^{-|y|}\, dy$$

$$= \frac{1}{2}\left[\int_{-\infty}^{0} (\alpha+\beta y)^r e^{y}\, dy + \int_{0}^{\infty} (\alpha+\beta y)^r e^{-y}\, dy\right]$$

For first integral, change y to $-y$

$$= \frac{1}{2}\left[\int_{0}^{\infty} (\alpha-\beta y)^r e^{-y}\, dy + \int_{0}^{\infty} (\alpha+\beta y)^r e^{-y}\, dy\right]$$

$$= \frac{1}{2}\int_{0}^{\infty} \{(\alpha-\beta y)^r + (a+\beta y)^r\} e^{-y}\, dy$$

$$= \int_{0}^{\infty} \{\alpha^r + {}^rc_2\, \alpha^{r-2} \beta^2 y^2 + {}^rc_4\, \alpha^{r-4} \beta^4 y^4 + \ldots\} e^{-y}\, dy$$

$$= \alpha^r + {}^rc_2\, \alpha^{r-2} \beta^2\, 2! + {}^rc_4\, \alpha^{r-4} \beta^4\, 4! + \ldots\ldots$$

$$= \alpha^r + 2!\, {}^rc_2\, \alpha^{r-2} \beta^2 + 4!\, {}^rc_4\, \alpha^{r-4} \beta^4 + \ldots\ldots$$

Put $r = 1$

$\therefore \quad \bar{x} = \mu'_1(0) = \alpha$

Put $r = 2$

$\mu'_2(0) = \alpha^2 \cdot$

$= \alpha^2 \cdot$

$\therefore \quad \text{var}(x) = \mu'_2$

$= \alpha^2 \cdot$

$= 2\beta^2$

12.7. Lognormal Distribution

Let x be a positive random variat

$y = \log$

Then, if y is normal variate, x is ca

lognormal distribution.

If y is $N(m, \sigma)$, dist. of y is

$dP = \frac{}{\sigma}$

$\therefore$ Dist. of x is given by

$dP = \frac{}{\sigma}$

$= \frac{}{\sigma}$

12.7.1. Moments, Mean and Variar

$\mu'_r(0)_{\text{of } x} = E($

$= E($

$= M$

$= e$

$\therefore \quad \bar{x} = \mu'_1(0) = e$

$\mu'_2(0) = e^2$

$\therefore \quad \text{var}(x) = \mu'$

$= e^2$

e^2

12.7.2. Theorem :

If $x_1, x_2, \ldots\ldots x_n$ be independent lo

σ_i then

$x = x_1$

is also a lognormal variate.

Proof : Let $y_i = \log x_i$

Then y_i is $N(m_i, \alpha_i)$, $\forall_i$.

Let $x = x_1$

$\therefore \quad \log x = \text{lo}$

$e \ldots\ldots\ldots^{-\alpha_n X}$

$_i)x$

$, -\infty < x < \infty$

dx

dy

$dy + \int_0^\infty (\alpha+\beta y)^r e^{-y}\, dy\Big]$

$dy + \int_0^\infty (\alpha+\beta y)^r e^{-y}\, dy\Big]$

$+ \beta y)^r\} e^{-y}\, dy$

$y^2 + {}^r c_4\, \alpha^{r-4} \beta^4 y^4 + \ldots\} e^{-y}\, dy$

${}^r c_4\, \alpha^{r-4} \beta^4\, 4! + \ldots\ldots$

$4!\, {}^r c_4\, \alpha^{r-4} \beta^4 + \ldots\ldots$

$$\begin{aligned}
\mu'_2(0) &= \alpha^2 + 2!^2 c_2 \beta^2 \\
&= \alpha^2 + 2\beta^2 \\
\therefore \quad \text{var}(x) &= \mu'_2(0) - \bar{x}^2 \\
&= \alpha^2 + 2\beta^2 - \alpha^2 \\
&= 2\beta^2
\end{aligned}$$

12.7. Lognormal Distribution

Let x be a positive random variate and

$$y = \log_e x.$$

Then, if y is normal variate, x is called **lognormal variate** and distribution of x is called **lognormal distribution.**

If y is $N(m, \sigma)$, dist. of y is

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-m}{\sigma}\right)^2} dy, \; -\infty < y < \infty.$$

$\therefore$ Dist. of x is given by

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - m}{\sigma}\right)^2} d(\log x)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - m}{\sigma}\right)^2} \frac{dx}{x}, \; x > 0.$$

12.7.1. Moments, Mean and Variance

$$\begin{aligned}
\mu'_r(0)_{\text{of } x} &= E(x^r) \\
&= E(e^{ry}) \\
&= M_0(r)_{\text{of } y} \\
&= e^{rm + \frac{1}{2} r^2 \sigma^2} \\
\therefore \quad \bar{x} = \mu'_1(0) &= e^{m + \frac{1}{2}\sigma^2} \\
\mu'_2(0) &= e^{2m + 2\sigma^2} \\
\therefore \quad \text{var}(x) &= \mu'_2(0) - \bar{x}^2 \\
&= e^{2m+2\sigma^2} - e^{2m+\sigma^2} \\
&= e^{2m+\sigma^2}\left(e^{\sigma^2} - 1\right).
\end{aligned}$$

12.7.2. Theorem :

If $x_1, x_2, \ldots\ldots x_n$ be independent lognormal variates such that $\log x_i$ has mean m_i and $s.d.$, σ_i then

$$x = x_1 x_2 \ldots\ldots x_n$$

is also a lognormal variate.

Proof : Let $y_i = \log x_i$

Then y_i is $N(m_i, \alpha_i)$, $\forall_i$.

Let $\quad x = x_1 x_2 \ldots\ldots x_n$

$\therefore \quad \log x = \log x_1 + \log x_2 + \ldots\ldots + \log x_n.$

$$= y_1 + y_2 \ldots\ldots + y_n$$

$\therefore$ By additive property of normal variates, $\log x$ is a $N.V.$ with mean

$$\sum_{i=1}^{n} m_i \text{ and } s.d. \sqrt{\sum_{i=1}^{n} \sigma_i^2}$$

$\therefore$ x is lognormal variate.

12.8. Cauchy Distribution

It is a continuous distribution with probability differential

$$dP = \frac{1}{\pi\lambda} \frac{1}{1+\left(\frac{x-\mu}{\lambda}\right)^2} dx, \quad -\infty < x < \infty, \ \lambda > 0 \qquad \ldots(12.8\text{-}1)$$

λ; μ are the parameters.

Substituting $\frac{x-\mu}{\lambda} = z$, (12.8-1) takes the form

$$dP = \frac{1}{\pi} \cdot \frac{dz}{1+z^2}, \quad -\infty < z < \infty \qquad \ldots(12.8\text{-}2)$$

This is standard cauchy distribution and z is called standard cauchy variate.
Distribution function of (12.8-1) is

$$F(x) = \frac{1}{\pi\lambda} \int_{-\infty}^{x} \frac{dx}{1+\left(\frac{x-\mu}{\lambda}\right)^2}$$

Put $\qquad \frac{x-\mu}{\lambda} = z$

$$= \frac{1}{\pi} \int_{-\infty}^{\frac{x-\mu}{\lambda}} \frac{dz}{1+z^2}$$

$$= \frac{1}{\pi} \left\{\tan^{-1} z\right\}_{-\infty}^{\frac{x-\mu}{\lambda}}$$

$$= \frac{1}{\pi} \tan^{-1} \frac{x-\mu}{\lambda} + \frac{1}{2}$$

For various constants of (12.8-2) sec. Ex. 9-28.

12.8.1. Moments

$$\bar{x} = E(x) = \frac{1}{\pi\lambda} \int_{-\infty}^{\infty} \frac{x\,dx}{1+\left(\frac{x-\mu}{\lambda}\right)^2}$$

Put $\qquad \frac{x-\mu}{\lambda} = z$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(\lambda z + \mu)\,dz}{1+z^2}$$

SPECIAL CONTINUOUS DISTRIBUTION

= λ

where I =

Now in general I does not exist

exists and is zero.

$\therefore$ In general, $\bar{x}$ does not e

(12.8-3) $\Rightarrow$ $\bar{x} = \mu$.

$\therefore$ μ_2 =

Put $\frac{x-\mu}{\lambda}$ =

which does not exist as the integral

$\therefore$ μ_2 does not exist. Simila

12.8.2. Median and Mode

From (12.8-1) eq. of prob. de

y =

Put $\frac{x-\mu}{\lambda}$ =

$\therefore$ y =

which is symmetrical about

$z = 0$ *i.e.*

$\therefore$ median is

Also obviously y is maximum

$\therefore$ Mode is at z =

i.e., x =

$\therefore$ Mean, mode and media

g x is a $N.V.$ with mean

fferential

$x, \; -\infty < x < \infty, \; \lambda > 0$ ...(12.8-1)

$z < \infty$ ...(12.8-2)

d standard cauchy variate.

$$= \lambda \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{z\,dz}{1+z^2} + \mu \cdot \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dz}{1+z^2}$$

$$= \frac{\lambda}{\pi} I + \frac{\mu}{\pi} \left\{\tan^{-1} z\right\}_{-\infty}^{\infty}$$

$$= \frac{\lambda}{\pi} I + \mu \qquad (12.8\text{-}3)$$

where $$I = \int_{-\infty}^{\infty} \frac{z\,dz}{1+z^2}$$

Now in general I does not exist but its principal value

$$\underset{t\to\infty}{Lin} \int_{-t}^{t} \frac{z}{1+z^2}\,dz$$

exists and is zero.

$\therefore$ In general, $\bar{x}$ does not exit but if the principal value of I is taken

(12.8-3) $\Rightarrow$ $\bar{x} = \mu.$

$$\therefore \quad \mu_2 = E(x-\mu)^2$$

$$= \frac{1}{\pi\lambda} \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{1+\left(\frac{x-\mu}{\lambda}\right)^2}\,dx$$

Put $$\frac{x-\mu}{\lambda} = z$$

$$= \frac{\lambda^2}{\pi} \int_{-\infty}^{\infty} \frac{z^2}{1+z^2}\,dz$$

which does not exist as the integral on right-hand side is not convergent.

$\therefore$ μ_2 does not exist. Similarly μ_r $(r > 2)$ does not exist.

12.8.2. Median and Mode

From (12.8-1) eq. of prob. density curve is

$$y = \frac{1}{\pi\lambda} \cdot \frac{1}{1+\left(\frac{x-\mu}{\lambda}\right)^2}$$

Put $$\frac{x-\mu}{\lambda} = z$$

$$\therefore \quad y = \frac{1}{\pi\lambda} \cdot \frac{1}{1+z^2}$$

which is symmetrical about

$z = 0$ *i.e.* $x = \mu$

$\therefore$ median is $x = \mu$

Also obviously y is maximum for $z = 0$

$\therefore$ Mode is at $z = 0$

i.e., $x = \mu$

$\therefore$ Mean, mode and median coincide.

12.8.3. Characteristic Function (c.f.)

From (12.8-2) *c.f.* is

$$\phi(t) = E(e^{itz})$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} e^{itz}\frac{1}{1+z^2}\,dz \qquad \text{changing } z \text{ to } -y$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-ity}\frac{1}{1+y^2}\,dy$$

Consider

$$I = \int_{-\infty}^{\infty} e^{ity}\,.\,e^{-|y|}\,dy$$

$$= \int_{-\infty}^{\infty} \{\cos(ty) + i\sin(ty)\}\,e^{-|y|}\,dy$$

$$= 2\int_{0}^{\infty} \cos(ty)\,e^{-y}\,dy$$

$$= 2\left[\left|\frac{\sin ty}{t}\cdot e^{-y}\right|_0^{\infty} + \frac{1}{t}\int_0^{\infty} e^{-y}\sin(ty)\,dy\right]$$

$$= \frac{2}{t}\left[\left|\frac{\cos ty\, e^{-y}}{-t}\right|_0^{\infty} - \frac{1}{t}\int_0^{\infty} e^{-y}\cos(ty)\,dy\right]$$

$$= \frac{2}{t^2} - \frac{2}{t^2}\int_0^{\infty} e^{-y}\cos(ty)\,dy$$

$$= \frac{2}{t^2} - \frac{I}{t^2}$$

$$\therefore \qquad I = \frac{2}{1+t^2}$$

$\therefore$ By inversion theorem, 8.5.4.

$$e^{-|y|} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\,.\,\frac{2}{1+t^2}\,dt$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-ity}\,.\,\frac{1}{1+t^2}\,dt$$

$$e^{-|t|} = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-ity}\,.\,\frac{dy}{1+y^2} = \phi(t) \qquad \text{interchanging } t \text{ and } y$$

Now, for (12.8-1) c.f. is

$$\phi(t) = E(e^{itx})$$

= -

Put $\frac{x-\mu}{\lambda} = z$

= -

= e

= (

12.8.4. Additive Property

Let $x_1, x_2, \ldots, x_n$ be independer

(λ_n, μ_n) respectively.

Then $\phi_{x_i}(t) = e$

Let $x = x$

$\phi_x(t) = E$

= E

= E

= ϕ

= (

$\therefore$ By uniqueness theorem o

x is a Cauchy Variate

with parameters $(\lambda_1 + \ldots + \lambda_n, \mu_1$

12.9. Truncated Distributions

Let x be a random variate with

Then density f^n of x truncated

$\overline{F(b}$

Ex. 12-2. *Let x be normally d density of x on the left at 'a' and on distribution. Furthermore, if a = μ – is m.*

Sol. Dist. of x is

$dP = -$

$\therefore \quad f(x) = -$

and $\quad F(x) = -$

changing z to $-y$

ty

$(ty)\} e^{-|y|} dy$

$\frac{1}{t} \int_0^\infty e^{-y} \sin(ty)\, dy \Big]$

$\frac{1}{t} \int_0^\infty e^{-y} \cos(ty)\, dy \Big]$

$ty) dy$

dt

t

$\phi(t)$ interchanging t and y

$$= \frac{1}{\pi\lambda} \int_{-\infty}^{\infty} \frac{e^{itx}\, dx}{1 + \left(\frac{x-\mu}{\lambda}\right)^2}$$

Put $\quad \dfrac{x-\mu}{\lambda} = z$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} e^{it(\mu+\lambda z)} \frac{dz}{1+z^2}$$

$$= e^{it\mu} \cdot \frac{1}{\pi} \int_{-\infty}^{\infty} e^{i(t\lambda)z} \cdot \frac{dz}{1+z^2}$$

$$= e^{it\mu}\, e^{-|t\lambda|} = e^{it\mu - \lambda|t|}.$$

12.8.4. Additive Property

Let $x_1, x_2, \ldots x_n$ be independent cauchy variates with parameters $(\lambda_1, \mu_1), (\lambda_2, \mu_2) \ldots (\lambda_n, \mu_n)$ respectively.

Then $\quad \phi_{x_j}(t) = e^{it\mu_j - \lambda_j|t|}$

Let $\quad x = x_1 + x_2 + \ldots + x_2$

$$\phi_x(t) = E\{e^{itx}\}$$

$$= E\{e^{itx_1} \cdot e^{itx_2} \ldots e^{itx_n}\}$$

$$= E(e^{itx_1}) \cdot E(e^{itx_2}) \ldots E(e^{itx_n})$$

$$= \phi_{x_1}(t)\, \phi_{x_2}(t) \ldots \phi_{x_n}(t)$$

$$= e^{it(\mu_1 + \ldots + \mu_n) - (\lambda_1 + \ldots + \lambda_n)|t|}$$

$\therefore$ By uniqueness theorem of c.f.s.

x is a Cauchy Variate

with parameters $(\lambda_1 + \ldots + \lambda_n, \quad \mu_1 + \ldots + \mu_n)$

12.9. Truncated Distributions

Let x be a random variate with density f^n $f(x)$ and cumulative distribution f^n $F(x)$.

Then density f^n of x truncated on the left at $x = a$ and on the right at $x = b$ is

$$\frac{f(x)}{F(b) - F(a)}, \quad a \le x \le b.$$

Ex. 12-2. *Let x be normally distributed with mean m and variance σ^2. Truncate the density of x on the left at 'a' and on the right at 'b' and calculate the mean of the truncated distribution. Furthermore, if $a = \mu - \sigma$ and $b = \mu + \sigma$, then mean of the truncated distribution is m.*

Sol. Dist. of x is

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx \quad -\infty < x < \infty$$

$$\therefore \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

and

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$\therefore \quad F(b)-F(a) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{b} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx - \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{a} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{a}^{b} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

Put $\qquad \dfrac{x-m}{\sigma} = y$

$$\therefore \quad F(b)-F(a) = \frac{1}{\sqrt{2\pi}}\int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} e^{-\frac{1}{2}y^2}\, dy \qquad \ldots(1)$$

$\therefore$ Density f^n of truncated distribution is

$$\frac{f(x)}{F(b)-F(a)} = \frac{\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}}{F(b)-F(a)}$$

and $\qquad a \le x \le b.$

$\therefore$ mean of the truncated dist. is

$$\bar{x} = \frac{1}{F(b)-F(a)}\,\frac{1}{\sigma\sqrt{2\pi}}\int_{a}^{b} x\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

Put $\quad \dfrac{x-m}{\sigma} = y$

$$= \frac{1}{F(b)-F(a)}\,\frac{1}{\sqrt{2\pi}}\int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} (m+\sigma y)\,.\, e^{-\frac{1}{2}y^2}\, dy$$

$$= \frac{1}{F(b)-F(a)}\left[m\,\frac{1}{\sqrt{2\pi}}\int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} e^{-\frac{1}{2}y^2}\, dy + \frac{\sigma}{\sqrt{2\pi}}\int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} y\, e^{-\frac{1}{2}y^2}\, dy\right]$$

$$= m + \frac{1}{\sqrt{2\pi}}\,\frac{\sigma}{F(b)-F(a)}\left\{-e^{-\frac{1}{2}y^2}\right\}_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}}$$

$$= m + \frac{1}{\sqrt{2\pi}}\,\frac{\sigma}{F(b)-F(a)}\left\{e^{-\frac{1}{2}\left(\frac{a-m}{\sigma}\right)^2} - e^{-\frac{1}{2}\left(\frac{b-m}{\sigma}\right)^2}\right\}$$

which lies between a and b

If $a = m-\sigma$ and $b = m+\sigma$ then

$$m-a = b-m = \sigma$$

$$\therefore \quad \bar{x} = m.$$

Ex. 12-3. *If x is a $N(m, \sigma)$, then*

$z = \frac{1}{2}$

is a $\gamma\left(\frac{1}{2}\right)$.

Sol. Distribution of x is

$dP = \frac{}{\sigma}$

Put $\quad z = \frac{1}{2}$

$\Rightarrow \quad x = m$

$\therefore \quad dx = \sqrt{}$

$\therefore$ Distribution of z is

$\therefore \quad dP = c.$

where c is constant to be obtained s.

$$\int_0^\infty dP = 1$$

$\therefore \quad c$ is given by

$$c\int_0^\infty e^{-z} z^{-1/2}\, dz = 1$$

i.e., $\quad c\Gamma\left(\frac{1}{2}\right) = 1$

i.e., $\quad c =$

$\therefore$ Distribution of z is

$dP =$

which implies that z is a $\gamma\left(\frac{1}{2}\right)$.

Ex. 12-4. *If x is a $\gamma(\lambda)$, find E(*

variate.

Sol. Distribution of x is

$dP =$

Ex. 12-3. *If x is a N(m, σ), then*

$$z = \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2$$

is a $y\left(\frac{1}{2}\right)$.

Sol. Distribution of x is

$$dP = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx \qquad -\infty < x < \infty$$

Put

$$z = \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2$$

$$\Rightarrow \qquad x = m + \sigma\sqrt{2z}$$

$$\therefore \qquad dx = \sqrt{2}\,\sigma\, \frac{dz}{2\sqrt{z}}$$

$\therefore$ Distribution of z is

$$\therefore \qquad dP = c.e^{-z}\,.\,z^{-1/2}\,dz,\ 0 < z < \infty$$

where c is constant to be obtained s.t.

$$\int_0^\infty dP = 1$$

$\therefore$ c is given by

$$c\int_0^\infty e^{-z}\,z^{-1/2}\,dz = 1$$

i.e.,

$$c\Gamma\left(\frac{1}{2}\right) = 1$$

i.e.,

$$c = \frac{1}{\Gamma\left(\frac{1}{2}\right)}$$

$\therefore$ Distribution of z is

$$dP = \frac{1}{\Gamma\left(\frac{1}{2}\right)}\, e^{-z}\,z^{-1/2}\,dz, \quad 0 < z < \infty$$

which implies that z is a $\gamma\left(\frac{1}{2}\right)$.

Ex. 12-4. *If x is a γ(λ), find* $E\left(\sqrt{x}\right)$. *Deduce mean deviation about mean for a normal variate.*

Sol. Distribution of x is

$$dP = \frac{1}{\Gamma(\lambda)}\, e^{-x}\,.\,x^{\lambda-1}\,dx, \quad 0 < x < \infty.$$

$$E\left(\sqrt{x}\right) = \frac{1}{\Gamma(\lambda)}\int_0^\infty \sqrt{x}\,.e^{-x}.x^{\lambda-1}\,dx$$

$$= \frac{1}{\Gamma(\lambda)}\int_0^\infty e^{-x}.x^{\left(\lambda+\frac{1}{2}\right)-1}\,dx$$

$$\frac{1}{\Gamma(\lambda)}\,\Gamma\left(\lambda+\frac{1}{2}\right)$$

Deduction. Let x be a $N(m, \sigma)$.

Then $$z = \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2 \text{ is a } \gamma\left(\frac{1}{2}\right)$$

Now $$|x-m| = \sigma\sqrt{2z}$$

$\therefore$ Mean deviation about mean

$$= E\,|x-m|$$

$$= \sigma\sqrt{2}\,.E\left(\sqrt{z}\right)$$

$$= \sigma\sqrt{2}\,.\,\frac{\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)} \qquad \left(\because\ \lambda = \frac{1}{2}\right)$$

$$= \sigma\sqrt{\frac{2}{\pi}}.$$

Ex. 12-5. *If x and y are independent gamma variates, find the distribution of*

(*i*) $x+y$ (*ii*) $\dfrac{x}{x+y}$ (*iii*) $\dfrac{x}{y}$.

Sol. Let x and y be gamma variates with parameters λ and μ respectively. Then distribution of x and y are

$$dP = \frac{1}{\Gamma(\lambda)}\,e^{-x}.x^{\lambda-1}\,dx, \quad 0<x<\infty$$

$$dP = \frac{1}{\Gamma(\mu)}\,e^{-y}.y^{\mu-1}\,dy, \quad 0<y<\infty$$

Since x and y are independent, joint distribution of x and y is

$$dP = \frac{1}{\Gamma(\lambda)\Gamma(\mu)}\,e^{-(x+y)}.\,x^{\lambda-1}\,y^{\mu-1}\,dx\cdot dy$$

Put $$u = x+y, \quad v = \frac{x}{x+y}$$

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\[2mm] \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

Now $x = uv$

$\therefore$ Joint distribution of u and

$\Rightarrow$ u and v are independent var

To find dist. of $\frac{x}{y}$; proceed as be

Put

Now

$\therefore$ Joint distribution of u and v i

$x^{\lambda-1}\, dx$

$\left(\frac{1}{2}\right)^{-1} dx$

$\left(\frac{1}{2}\right)$

$\left(\because \quad \lambda = \frac{1}{2}\right)$

ates, find the distribution of $\frac{x}{y}$.

ers λ and μ respectively.

$, \quad 0 < x < \infty$

$, \quad 0 < y < \infty$

f x and y is

$x^{\lambda-1}\, y^{\mu-1}\, dx \cdot dy$

$$= \begin{vmatrix} 1 & 1 \\ \dfrac{y}{(x+y)^2} & \dfrac{-x}{(x+y)^2} \end{vmatrix}$$

$$= -\frac{(x+y)}{(x+y)^2} = -\frac{1}{x+y} = -\frac{1}{u}$$

Now $\qquad x = uv, \quad y = u(1-v)$

$\therefore$ Joint distribution of u and v is

$$dP = \frac{1}{\Gamma(\lambda)\Gamma(\mu)} e^{-u} (uv)^{\lambda-1} \{u(1-v)\}^{\mu-1} u.\, du\, dv$$

$$= \frac{1}{\Gamma(\lambda)\Gamma(\mu)} e^{-u} (u)^{\lambda+\mu-1} . v^{\lambda-1} (1-v)^{\mu-1}\, du\, dv$$

$$= \left\{\frac{1}{\Gamma(\lambda+\mu)} e^{-u} . u^{\lambda+\mu-1} . du\right\} \left\{\frac{1}{\beta(\lambda,\mu)} v^{\lambda-1} (1-v)^{\mu-1}\, dv\right\}$$

$\Rightarrow$ u and v are independent variates. u is a $\gamma(\lambda+\mu)$ variate and v is $\beta_1(\lambda, \mu)$ variate.

To find dist. of $\frac{x}{y}$; proceed as below :

Put $\qquad u = x+y, \; v = \frac{x}{y}$

$$\therefore \quad \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 \\ \dfrac{1}{y} & -\dfrac{x}{y^2} \end{vmatrix}$$

$$= -\frac{x+y}{y^2}$$

Now $\qquad x = \frac{uv}{1+v}, \; y = \frac{u}{1+v}$

$$\therefore \quad \frac{\partial(u,v)}{\partial(x,y)} = -\frac{(1+v)^2}{u}$$

$\therefore$ Joint distribution of u and v is

$$dP = \frac{1}{\Gamma(\lambda)\Gamma(\mu)} e^{-u} . \left(\frac{uv}{1+v}\right)^{\lambda-1} \left(\frac{u}{1+v}\right)^{\mu-1} du\, dv\, \frac{u}{(1+v)^2}$$

$$= \frac{1}{\Gamma(\lambda)\Gamma(\mu)} . e^{-u} . u^{\lambda+u-1} . \frac{v^{\lambda-1}}{(1+v)^{\lambda+u}}\, du\, dv$$

$$= \left\{\frac{1}{\Gamma(\lambda+\mu)} e^{-u} . u^{\lambda+u-1}\, du\right\} \left\{\frac{1}{\beta(\lambda,\mu)} \frac{v^{\lambda-1}}{(1+v)^{\lambda+u}}\, dv\right\}$$

$\Rightarrow$ u and v are independent variates

u is a $\gamma(\lambda + \mu)$ variate

and v is $\beta_2(\lambda, \mu)$ variate.

Ex. 12-6. *If x and y are two independent standard normal variates, find the distributions of (i) x^2 (ii) $x^2 + y^2$ (iii) $\dfrac{x^2}{y^2}$.*

Sol. (*i*) Put $u = x^2$

Then $\dfrac{1}{2} u$ is a $\gamma\left(\dfrac{1}{2}\right)$ variate.

$\therefore$ Dist. of $\dfrac{1}{2} u$ *is*

$$dP = \frac{1}{\Gamma\left(\frac{1}{2}\right)} e^{-\frac{u}{2}} \left(\frac{u}{2}\right)^{\frac{1}{2}-1} d\left(\frac{u}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} u^{-\frac{1}{2}} du$$

which gives the distribution of u.

(*ii*) $\dfrac{1}{2} x^2$ is a $\gamma\left(\dfrac{1}{2}\right)$ variate

$\dfrac{1}{2} y^2$ is a $\gamma\left(\dfrac{1}{2}\right)$ variate

$\therefore$ $\dfrac{1}{2}(x^2 + y^2)$ is a $\gamma\left(\dfrac{1}{2} + \dfrac{1}{2} = 1\right)$ variate.

$\therefore$ Dist. of $\dfrac{x^2 + y^2}{2}$ is

$$dP = \frac{1}{\Gamma(1)} e^{-\frac{1}{2}(x^2+y^2)} \left(\frac{x^2+y^2}{2}\right)^{1-1} d\left(\frac{x^2+y^2}{2}\right)$$

$$= \frac{1}{2} e^{-\frac{1}{2}\psi^2} d\psi^2$$

where $\psi^2 = x^2 + y^2$

(*iii*) $\dfrac{1}{2} x^2$ is a $\gamma\left(\dfrac{1}{2}\right)$ variate

$\dfrac{1}{2} y^2$ is a $\gamma\left(\dfrac{1}{2}\right)$ variate

$$\therefore \quad u = \frac{\frac{1}{2}x^2}{\frac{1}{2}y^2} = \frac{x^2}{y^2} \text{ is a } \beta_2\left(\frac{1}{2}, \frac{1}{2}\right).$$

Ex. 12-7. *If x and y are gamma variates with parameter λ and μ. Find the distributions of $x + y$ and $\dfrac{x - y}{x + y}$.*

Sol. Let $u = x + y$, $v = $ ...

$\therefore$ $x = $ ...

$\therefore$ Now distributions of x and ...

The joint distribution of x and y i...

The joint distribution of u and v ...

$\therefore$ u and v are independent and th...

Ex. 12.8. *Let x be a $\beta_1(\lambda, \mu)$, fi...*

Sol. Dist. of x is

l normal variates, find the distributions

$$^{-1} d\left(\frac{u}{2}\right)$$

$$\left(\frac{x^2+y^2}{2}\right)^{1-1} d\left(\frac{x^2+y^2}{2}\right)$$

$$\beta_2\left(\frac{1}{2}, \frac{1}{2}\right).$$

rameter λ and μ. Find the distributions

Sol. Let $\quad u = x + y,\ v = \dfrac{x-y}{x+y}$

$$\therefore \quad x = \frac{u(1+v)}{2},\quad y = \frac{u(1-v)}{2}$$

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{1+v}{2} & \dfrac{u}{2} \\ \dfrac{1-v}{2} & -\dfrac{u}{2} \end{vmatrix} = -\frac{u}{2}$$

$\therefore$ Now distributions of x and y are

$$dP = \frac{1}{\Gamma(\lambda)} e^{-x}, x^{\lambda-1}\, dx$$

$$dP = \frac{1}{\Gamma(\mu)} e^{-y}, y^{\mu-1}\, dy.$$

The joint distribution of x and y is

$$dP = \frac{1}{\Gamma(\lambda).\Gamma(\mu)} e^{-(x+y)} x^{\lambda-1} y^{\mu-1}\, dx\, dy.$$

The joint distribution of u and v is

$$dP = \frac{1}{\Gamma(\lambda).\Gamma(\mu)} e^{-u}.\left\{\frac{u(1+v)}{2}\right\}^{\lambda-1}\left\{\frac{u(1-v)}{2}\right\}^{\mu-1}.\, du\, dv\,.\,\frac{u}{2}$$

$$= \left\{\frac{1}{\Gamma(\lambda+\mu)} e^{-u}.\, u^{\lambda+\mu-1}\, du\right\}.$$

$$\left\{\frac{1}{2^{\lambda+\mu-1}\,\beta(\lambda,\mu)} (1+v)^{\lambda-1}(1-v)^{\mu-1}\, dv\right\}$$

$\therefore$ u and v are independent and their distributions are

$$dP = \frac{1}{\Gamma(\lambda+\mu)} e^{-u} u^{\lambda+\mu-1}\, du$$

$$dP = \frac{1}{2^{\lambda+\mu-1}\,\beta(\lambda,\mu)} (1+v)^{\lambda-1}(1-v)^{\mu-1}\, dv \text{ respectively.}$$

Ex. 12.8. *Let x be a* $\beta_1(\lambda, \mu)$, *find* $E(\sqrt{x})$.

Sol. Dist. of x is

$$dP = \frac{1}{\beta(\lambda,\mu)} x^{\lambda-1}(1-x)^{\mu-1}\, dx, \quad 0 < x < 1$$

$$E\left(\sqrt{x}\right) = \frac{1}{\beta(\lambda,\mu)} \int_0^1 x^{\frac{1}{2}} \cdot x^{\lambda-1} (1-x)^{\mu-1}\, dx$$

$$= \frac{1}{\beta(\lambda,\mu)} \int_0^1 x^{\lambda+\frac{1}{2}-1} (1-x)^{\mu-1}\, dx$$

$$= \frac{1}{\beta(\lambda,\mu)}\, \beta\left(\lambda+\frac{1}{2}, \mu\right)$$

$$= \frac{\Gamma\left(\lambda+\frac{1}{2}\right)\Gamma(\lambda+\mu)}{\Gamma(\lambda)\Gamma\left(\lambda+\mu+\frac{1}{2}\right)}.$$

Ex. 12-9. *If x and y be independent standard cauchy variates, find the dist. of xy.*

Sol. The joint dist. of x and y is

$$dP = \left\{\frac{1}{\pi}\,\frac{1}{1+x^2}\, dx\right\} \cdot \left\{\frac{1}{\pi} \cdot \frac{1}{1+y^2}\, dy\right\}$$

$$= \frac{1}{\pi^2} \cdot \frac{1}{(1+x^2)(1+y^2)}\, dx\, dy$$

Put $u = xy,\ v = y$

Both u, v vary from $-\infty$ to $+\infty$

$$\therefore \quad x = \frac{u}{v},\ y = v$$

$$\therefore \quad J = \begin{vmatrix} \frac{1}{v} & \frac{-u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v}$$

$\therefore$ Joint dist. of u, v is

$$dP = \frac{1}{\pi^2\left(1+\frac{u^2}{v^2}\right)(1+v^2)}\, \frac{1}{|v|}\, du\, dv$$

$$= \frac{1}{\pi^2}\, \frac{|v|}{(u^2+v^2)(1+v^2)}\, du\, dv$$

$\therefore$ Dist. of u is

$$dP = \frac{du}{\pi^2} \int_{-\infty}^{\infty} \frac{|v|\, dv}{(u^2+v^2)(1+v^2)}$$

$$= \frac{2du}{\pi^2} \int_0^{\infty} \frac{v\, dv}{(u^2+v^2)(1+v^2)}$$

Put $v^2 = t$

$$= \frac{du}{\pi^2} \int_0^{\infty} \frac{dt}{(u^2+t)(1+t)}$$

$$= \frac{du}{\pi^2}$$

$$= \frac{\quad}{\pi^2}$$

$$= \frac{\quad}{\pi^2}$$

$$= \frac{\quad}{\pi^2}$$

1. If $x_1, x_2, \ldots, x_n$ are gamma va
$\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$.

2. For a Beta distribution of first k
(*i*) A.M. > H.M.
(*ii*) $\log G = \frac{1}{\beta(\lambda,\mu)} \frac{\partial}{\partial\lambda} \beta(\lambda,$

3. If $X_1, X_2, \ldots X_n$ are independent
Y_n = Ma
Z_n = Mi
Then c.d.f.
$F_{Y_n}(y) = \{F$
$F_{Z_n}(y) = 1 -$
respectively.

Sol. (*i*) $F_{Y_n}(y) = P\{$
$= P\{$
$= \prod_{i=1}^{n}$
$= \prod_{i=1}^{n}$

(*ii*) $F_{Z_n}(y) = P\{$
$= 1 -$
$= 1 -$
$= 1 -$
$= 1 -$
$= 1 -$
$= 1 -$

$^{-1}(1-x)^{\mu-1}\,dx$

$^{1}(1-x)^{\mu-1}\,dx$

$\cdot,\mu)$

$\cdot\mu)$

$\left.\frac{1}{2}\right)$

auchy variates, find the dist. of xy.

$\left\{\frac{1}{\pi}\cdot\frac{1}{1+y^2}\,dy\right\}$

$\overline{y^2)}\,dx\,dy$

$\overline{-v^2)}\,\frac{1}{|v|}\,du\,dv$

$\overline{\cdot v^2)}\,du\,dv$

$\frac{|dv}{^2)(1+v^2)}$

$\frac{dv}{^2)(1+v^2)}$

$\overline{1+t)}$

$$= \frac{du}{\pi^2}\int_0^\infty \frac{1}{(1-u^2)}\left\{\frac{1}{u^2+t}-\frac{1}{1+t}\right\}dt$$

$$= \frac{du}{\pi^2(1-u^2)}\left|\log\left\{\frac{u^2+t}{1+t}\right\}\right|_0^\infty$$

$$= \frac{du}{\pi^2(1-u^2)}\left|-\log u^2\right|$$

$$= \frac{2du}{\pi^2(u^2-1)}\log|u|.$$

EXERCISES

1. If $x_1, x_2, \ldots, x_n$ are gamma variates each with parameter λ, find the distribution of $\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$.

2. For a Beta distribution of first kind, show that
   (*i*) A.M. > H.M.
   (*ii*) $\log G = \frac{1}{\beta(\lambda,\mu)}\frac{\partial}{\partial\lambda}\beta(\lambda,\mu)$.

3. If $X_1, X_2, \ldots X_n$ are independent random variables with same c.d.f. $F_X(.)$ let
$$Y_n = \text{Max}\{X_1, X_2, \ldots X_n\}$$
$$Z_n = \text{Min}\{X_1, X_2, \ldots X_n\}$$
Then c.d.f. of Y_n and Z_n are given by
$$F_{Y_n}(y) = \{F_X(y)\}^n$$
$$F_{Z_n}(y) = 1-\{1-F_X(y)\}^n$$
respectively.

Sol. (*i*)
$$\begin{aligned}F_{Y_n}(y) &= P\{Y_n \le y\}\\ &= P\{X_1 \le y;\ X_2 \le y_n, \ldots, X_n \le y\}\\ &= \prod_{i=1}^n P(X_i \le y) \qquad (\because\ X\text{'s are independent})\\ &= \prod_{i=1}^n F_X(y) = \{F_X(y)\}^n\end{aligned}$$

(*ii*)
$$\begin{aligned}F_{Z_n}(y) &= P\{Z_n \le y\}\\ &= 1-P(Z_n > y)\\ &= 1-P\{X_1 > y;\ X_2 > y; \ldots; X_N > y)\\ &= 1-\prod_{i=1}^n P(X_i > y)\\ &= 1-\prod_{i=1}^n \{1-P(X_i \le y)\}\\ &= 1-\prod_{i=1}^n \{1-F_X(y)\}\\ &= 1-\{1-F_X(y)\}^n.\end{aligned}$$

❑❑❑

13

Correlation Co-efficient and Linear Regression

13.1. Introduction

In bivariate distributions there are two variates x and y. If the change in one affects the change in other, the variables are said to be correlated. Otherwise they are said to be uncorrelated. If the increase (or decrease) in one results the increase (or decrease) in other, the correlation is said to be positive otherwise negative.

In bivariate distribution the data is of the form $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$. Each of these value pairs can be represented by a point in xy-plane. The resulting set of points is called a scatter diagram. From the scatter diagram, one can get a fairly good idea, though vague, about the correlation between the variables. If the points are very dense (*i.e.*, very close to each other) a fairly good amount of correlation is expected and if the points are widely scattered correlation is poor.

As a measure of degree of linear relationship between the variates, co-efficient of correlation is defined. This formula is referred to as product-moment formula for linear correlation. It, being due to Karl Pearson, is sometimes called Karl Pearson's correlation co-efficient.

13.2. Covariance

Covariance between two variates x and y is defined to be

$$E(x-\bar{x})(y-\bar{y}) \qquad \textit{(for prob. dist.)}$$

or

$$\frac{1}{N}\sum_{i=1}^{n} f_i (x_i-\bar{x})(y_i-\bar{y}) \qquad \textit{(for freq. dist.)}$$

where $\bar{x}$ and $\bar{y}$ are respective means and is denoted by 'cov (x, y)'.

Correlation co-efficient. *Correlation co-efficient between two variates x and y is defined to be*

$$\frac{\operatorname{cov}(x,y)}{(s.d.\, of\, x)(s.d.\, of\, y)}$$

and is denoted by r_{xy} or $\rho_{x,y}$

For freq. dist.

$$r_{xy} = \frac{\Sigma f(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma f(x-\bar{x})^2 \,.\, \Sigma f(y-\bar{y})^2}}$$

$$= \frac{\Sigma fxy - \frac{1}{N}(\Sigma fx)(\Sigma fy)}{\sqrt{\left\{\Sigma fx^2 - \frac{1}{N}(\Sigma fx)^2\right\}\left\{\Sigma fy^2 - \frac{1}{N}(\Sigma fy)^2\right\}}}$$

and for prob. dist.

$r_{xy} = \dots \sqrt{E \dots}$

$= \dots \sqrt{[\dots}$

13.2-1. Properties of Correlation C…

(*i*) *The correlation co-efficient is* …

Sol. Let x and y be two variates … correlation co-efficient between them …

The transformations correspondi…

$u = \frac{x - \dots}{h}$

where u and v are the variates to which… to the change of origin and h and k ar…

Now $x = a + \dots$

$\therefore \bar{x} = a + \dots$

where $\bar{u}, \bar{v}$ are expected values of u, …

Now $r_{xy} = \dots \sqrt{E \dots}$

$= \frac{hk \dots}{|hk \dots}$

according as h and k are of same or o…

$\therefore |r_{xy}| = |r_{uv} \dots$

(*ii*) *For two independent variate*…

Sol. Let x and y be two independe…

Now $\operatorname{cov}(x, y) = E\{\dots$

$= E\{\dots$

$= E(\dots$

$= E(\dots$

Since x and y are independent,

$E(xy) = E(x \dots$

$\therefore \operatorname{cov}(x, y) = \bar{x} \cdot \dots$

$\therefore r_{xy} = 0.$

Converse. It is not necessary that … consider the following example :

x :	– 3	–2
$y = x^2$:	9	4
xy :	– 27	– 8

Here $\Sigma x = 0 = \Sigma xy$

$$r_{xy} = \frac{E(x-\bar{x})(y-\bar{y})}{\sqrt{E(x-\bar{x})^2 \, E(y-\bar{y})^2}}$$

$$= \frac{E(xy) - E(x)E(y)}{\sqrt{[E(x)^2 - \{E(x)\}^2][E(y)^2 - \{E(y)\}^2]}}.$$

13.2-1. Properties of Correlation Co-efficient

(*i*) *The correlation co-efficient is numerically independent of origin and scale.*

Sol. Let x and y be two variates with expected values $\bar{x}$ and $\bar{y}$ respectively and r the correlation co-efficient between them.

The transformations corresponding to change of origin and scale are

$$u = \frac{x-a}{h} \quad \text{and} \quad v = \frac{y-b}{k}$$

where u and v are the variates to which x and y transform, a and b are constants corresponding to the change of origin and h and k are constants corresponding to change of scale.

Now $\quad x = a + uh \quad$ and $\quad y = b + kv$

$\therefore \quad \bar{x} = a + \bar{u}h$ and $\bar{y} = b + k\bar{v}$

where $\bar{u}$, $\bar{v}$ are expected values of u, v respectively.

Now
$$r_{xy} = \frac{E\{(x-\bar{x})(y-\bar{y})}{\sqrt{E(x-\bar{x})^2 \, E(y-\bar{y})^2}} = \frac{hkE\{(u-\bar{u})(v-\bar{v})}{\sqrt{h^2 k^2 \, E(u-\bar{u})^2 \, . \, E(v-\bar{v})^2}}$$

$$= \frac{hk}{|hk|} \, r_{uv} = \pm \, r_{uv}$$

according as h and k are of same or opposite signs.

$$\therefore \qquad |r_{xy}| = |r_{uv}|.$$

(*ii*) *For two independent variates correlation co-efficient is zero.*

Sol. Let x and y be two independent variates with expected values $\bar{x}$ and $\bar{y}$ respectively.

Now
$$\begin{aligned} \text{cov}(x, y) &= E\{(x - \bar{x})(y - \bar{y})\} \\ &= E\{xy - \bar{x}y - \bar{y}x + \bar{x} \cdot \bar{y}\} \\ &= E(xy) - \bar{x}\,E(y) - \bar{y}\,E(x) + \bar{x} \cdot \bar{y} \\ &= E(xy) - \bar{x} \cdot \bar{y} - \bar{x} \cdot \bar{y} + \bar{x} \cdot \bar{y} = E(xy) - \bar{x} \cdot \bar{y} \end{aligned}$$

Since x and y are independent,

$$E(xy) = E(x)\,E(y) = \bar{x} \cdot \bar{y}$$

$$\therefore \quad \text{cov}(x, y) = \bar{x} \cdot \bar{y} - \bar{x} \cdot \bar{y} = 0$$

$$\therefore \quad r_{xy} = 0.$$

Converse. It is not necessary that, if $r = 0$, the variates are independent. To observe this consider the following example :

x :	-3	-2	-1	0	1	2	3
$y = x^2$:	9	4	1	0	1	4	9
xy :	-27	-8	-1	0	1	8	27

Here $\Sigma x = 0 = \Sigma xy$

$$\therefore \qquad r_{xy} = \frac{\Sigma xy - \frac{1}{N}(\Sigma x)(\Sigma y)}{\sqrt{(\Sigma x^2) - \frac{1}{N}(\Sigma x)^2}\,.\,\sqrt{\Sigma y^2 - \frac{1}{N}(\Sigma y)^2}} = 0$$

(*iii*) *The correlation coefficient between linearly related variables is '+ 1' or '– 1'.*

Sol. Let x and y be the variates related by the equation $y = mx + c$ and $\bar{x}, \bar{y}$ be their expected values.

Then $$\bar{y} = m\bar{x} + c$$

Now $$r_{xy} = \frac{E(x-\bar{x})(y-\bar{y})}{E(x-\bar{x})^2\, E(y-\bar{y})^2} = \frac{mE(x-\bar{x})^2}{\sqrt{m^2}\,.E(x-\bar{x})}$$

$$= \frac{m}{|m|} = \pm 1$$

according as m is positive or negative.

(*iv*) *The correlation co-efficient cannot numerically exceed unity.*

Sol. Let x and y be the variates with expected values $\bar{x}$ and $\bar{y}$ respectively.

Let $$x' = x - \bar{x} \text{ and } y' = y - \bar{y}$$

Now for any real constant 'a', $(ax' - y')^2 \geq 0$

Since probabilities are non-negative and the sum of the non-negative quantities is non-negative,

$$E(ax' - y')^2 = \sum_i p_i (ax'_i - y'_i)^2 \geq 0$$

$$\therefore \quad a^2 E(x'^2) + E(y'^2) - 2aE(x'y') \geq 0$$

Put $$a = \frac{E(x'y')}{E(x'^2)}$$

Then $$E(y'^2) \geq \frac{\{E(x'y')\}^2}{E(x'^2)}$$

i.e., $$1 \geq \left\{\frac{E(x'y')}{E(x'^2)\, E(y')^2}\right\}^2 = \left\{\frac{E(x-\bar{x})(\bar{x}-\bar{y})}{\sqrt{E(x-\bar{x})^2\, E(y-\bar{y})^2}}\right\}^2 = r_{xy}^2$$

$$\therefore \quad |r_{xy}| \leq 1 \text{ or } -1 \leq r_{xy} \leq 1.$$

Ex. 13-1. *Show that, if x', y' are the deviations of the variables x, y from their means,*

$$r = 1 - \frac{1}{2N}\sum_i f_i \left(\frac{x'_i}{\sigma_x} - \frac{y'_i}{\sigma_y}\right)^2 \quad \text{(Symbols have their usual meanings)}$$

and $$r = -1 + \frac{1}{2N}\sum_i f_i \left(\frac{x'_i}{\sigma_x} + \frac{y'_i}{\sigma_y}\right)^2$$

Deduce that $-1 \leq r \leq 1$.

Sol. Consider $1 - \frac{1}{2N}\sum_i f_i$ ...

Similarly second result can be pr...

Ex. 13-2. *Show that the co-effici... expressed in the form*

where $\bar{x}$, $\bar{y}$ are A. Ms. and σ_x, σ_y are ...

Sol. By def.

Note. *For the numerical data r... new variates u and v defined by*

where a, b, h and k are constants to b... using the fact $r_{uv} = \pm r_{xy}$.

Sol. Consider $1-\dfrac{1}{2N}\sum_i f_i\left(\dfrac{x_i'}{\sigma_x}-\dfrac{y_i'}{\sigma_y}\right)^2$

$$= 1-\frac{1}{2N}\sum_i f_i\left(\frac{x_i'^2}{\sigma_x^2}+\frac{y_i'^2}{\sigma_y^2}-\frac{2x_i'y_i'}{\sigma_x\sigma_y}\right)^2$$

$$= 1-\frac{1}{2\sigma_x^2}\left(\frac{1}{N}\sum_i f_i x_i'^2\right)-\frac{1}{2\sigma_y^2}\left(\frac{1}{N}\sum f_i y_i'^2\right)$$

$$+\frac{1}{\sigma_x\sigma_y}\left(\frac{1}{N}\sum_i f_i x_i'\ y_i'\right)$$

$$= 1-\frac{1}{2\sigma_x^2}\cdot\sigma_x^2-\frac{1}{2\sigma_y^2}\cdot\sigma_y^2+\frac{\text{cov}(x,y)}{\sigma_x\sigma_y}=r$$

Similarly second result can be proved.

Now $$1-r = \frac{1}{2N}\sum_i f_i\left(\frac{x_i'}{\sigma_x}-\frac{y_i'}{\sigma_y}\right)^2 \ge 0 \ \Rightarrow\ 1\ge r$$

and $$1+r = \frac{1}{2N}\sum_i f_i\left(\frac{x_i'}{\sigma_x}+\frac{y_i'}{\sigma_y}\right)^2 \ge 0 \ \Rightarrow\ r\ge -1$$

$\therefore$ $$-1\le r\le 1.$$

Ex. 13-2. *Show that the co-efficient of correlation between two variates x and y may be expressed in the form*

$$\frac{1}{\sigma_x\sigma_y}\left(\frac{1}{N}\sum xy-\bar{x}\cdot\bar{y}\right)$$

where $\bar{x}, \bar{y}$ are A. Ms. and σ_x, σ_y are s.d.'s.

Sol. By def.

$$r_{xy} = \frac{\text{cov}(x,y)}{(s.d.\text{ of }x)(s.d.\text{ of }y)} = \frac{\frac{1}{n}\sum(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}$$

$$= \frac{1}{n\,\sigma_x\sigma_y}\left\{\sum(xy-\bar{x}y-\bar{y}x+\bar{x}\cdot\bar{y}\right\}$$

$$= \frac{1}{n\,\sigma_x\sigma_y}\left\{\sum xy-n\bar{x}\cdot\bar{y}-n\bar{x}\cdot\bar{y}+n\bar{x}\cdot\bar{y}\right\}$$

$$= \frac{1}{\sigma_x\sigma_y}\left\{\frac{1}{n}\sum xy-\bar{x}\cdot\bar{y}\right\}$$

Note. *For the numerical data r_{xy} is calculated by changing the variates x and y to the new variates u and v defined by*

$$u = \frac{x-a}{h}\ \textit{and}\ v=\frac{y-b}{k}$$

where a, b, h and k are constants to be chosen suitably so as to simplify the calculations and using the fact $r_{uv}=\pm\, r_{xy}$.

Ex. 13-3. *The ages (x) and systolic blood pressures (y) of 12 women are given below :*

Ages in years (x)	*Blood pressure (y)*
56	147
42	125
72	160
36	118
63	149
47	128
55	150
49	145
38	115
42	140
68	152
60	155

Calculate the correlation co-efficient between x and v.

Sol. Define the new variates

$u = x - 52$, and $v = y - 140$.

Then we have the table of values as below :

x	y	u	v	u^2	v^2	uv
56	147	4	7	16	49	28
42	125	– 10	– 15	100	225	150
72	160	20	20	400	400	400
36	118	– 16	– 22	256	484	352
63	149	11	9	121	81	99
47	128	– 5	– 12	25	144	60
55	150	3	10	9	100	30
49	145	– 3	5	9	25	– 15
38	115	– 14	– 25	196	625	350
42	140	– 10	0	100	0	0
68	152	16	12	256	144	192
60	155	8	15	64	225	120
		4	4	1552	2502	1766

$$r = \frac{1766 - \frac{1}{12}(4)(4)}{\sqrt{1552 - \frac{1}{12}(4)^2}\ \sqrt{2502 - \frac{1}{12}(4)^2}} \approx 0{\cdot}896.$$

Ex. 13-4. *Find the co-efficient*

$y \rightarrow$ / $x \downarrow$	67	72
92	—	—
87	—	—
82	4	4
77	3	3
72	2	3
67	3	2
62	1	—

Let $u = \frac{x}{}$

and $v = \frac{y}{}$

Sol. Calcula

$v \rightarrow$ / u	– 3	– 2	– 1	0
3	—	—	—	**0** 1
2	—	—	**– 2** 1	**0** 3
1	**– 12** 4	**– 8** 4	**– 6** 6	**0** 4
0	**0** 3	**0** 3	**0** 7	**0** 6
– 1	**6** 2	**6** 3	**5** 5	**0** 6
– 2	**18** 3	**8** 2	—	—
– 3	**9** 1	—	—	—
Total f	13	12	19	2(
fv	– 39	– 24	– 19	(
fv^2	117	48	19	(
fuv	21	6	– 3	(

(y) *of* 12 *women are given below* :

od pressure (y)

147
125
160
118
149
128
150
145
115
140
152
155

$d\,v$.

u^2	v^2	uv
16	49	28
100	225	150
400	400	400
256	484	352
121	81	99
25	144	60
9	100	30
9	25	– 15
196	625	350
100	0	0
256	144	192
64	225	120
1552	2502	1766

$$(4)(4) \quad \approx 0{\cdot}896.$$

$$2502-\frac{1}{12}(4)^2$$

Ex. 13-4. *Find the co-efficient of correlation for the following table :*

$y \rightarrow$ / $x \downarrow$	67	72	77	82	87	92	97
92	—	—	—	1	2	3	1
87	—	—	1	3	8	1	5
82	4	4	6	4	9	1	—
77	3	3	7	6	4	—	—
72	2	3	5	6	1	1	—
67	3	2	—	—	—	—	—
62	1	—	—	—	—	—	—

Let $$u = \frac{x-77}{5}$$

and $$v = \frac{y-82}{5}$$

Sol. Calculation of Co-eff. of Correlation.

$v \rightarrow$ / u	– 3	– 2	–1	0	1	2	3	Total f	fu	fu^2	fuv
3	—	—	—	**0** 1	**6** 2	**18** 3	**9** 1	7	21	63	33
2	—	—	**– 2** 1	**0** 3	**16** 8	**4** 1	**30** 5	18	36	72	48
1	– 12 4	– 8 4	– 6 6	**0** 4	**9** 9	**2** 1	—	28	28	28	– 15
0	**0** 3	**0** 3	**0** 7	**0** 6	**0** 4	—	—	23	0	0	0
– 1	**6** 2	**6** 3	**5** 5	**0** 6	– 1 1	– 2 1	—	18	– 18	18	14
– 2	**18** 3	**8** 2	—	—	—	—	—	5	– 10	20	26
– 3	**9** 1	—	—	—	—	—	—	1	– 3	9	9
Total f	13	12	19	20	24	6	6	100	54	210	115
fv	– 39	– 24	– 19	0	24	12	18	– 28			
fv^2	117	48	19	0	24	24	54	286			
fuv	21	6	– 3	0	30	22	39	115			

$$r = \frac{115 - \frac{1}{100}(54)(-28)}{\sqrt{210 - \frac{1}{100}(54)^2} \cdot \sqrt{286 - \frac{1}{100}(-28)^2}} \approx 0{\cdot}58.$$

Ex. 13-5. *A computer while calculating r_{xy} from 25 pairs of observation obtained the following constants :*

$$n = 25,\ \Sigma x = 125,\ \Sigma x^2 = 650,\ \Sigma y = 100,\ \Sigma y^2 = 460,\ \Sigma xy = 508.$$

A recheck showed that he had copied down two pairs (6, 14), (8, 6) *while the correct values were* (8, 12), (6, 8). *Obtain the correct value of the correlation co-efficient.*

Sol. $n = 25$

Correct Value of $\Sigma x = 125 - 6 - 8 + 8 + 6 = 125$

Correct Value of $\Sigma x^2 = 650 - 36 - 64 + 64 + 36 = 650$

Correct Value of $\Sigma y = 100 - 14 - 6 + 12 + 8 = 100.$

Correct Value of $\Sigma y^2 = 460 - 196 - 36 + 144 + 64 = 436$

Corrected Value of $\Sigma xy = 508 - 84 - 48 + 96 + 48 = 520$

$\therefore$ Correct value of co-efficient of correlation

$$= \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\ \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{(25)(520) - (125)(100)}{\sqrt{(25)(650) - (125)^2}\ \sqrt{(25)(436) - (100)^2}}$$

$$= \frac{520 - 500}{\sqrt{650 - 625}\ \sqrt{436 - 400}} = \frac{20}{5{\cdot}6} = \frac{2}{3}.$$

Ex. 13-6. *The following table gives the number of students having different hts. and wts :*

wts. in lbs.

hts. in inches	80 – 90	90 – 100	100 – 110	110 – 120	120 – 130	Total
50 – 55	1	3	7	5	2	18
55 – 60	2	4	10	7	4	27
60 – 65	1	5	12	10	7	35
65 – 70	0	3	8	6	3	20
Total	4	15	37	28	16	100

Calculate co-efficient of correlation between hts. and wts.

Sol. Let x and y be the variates for the wts. and hts. respectively.

Calculation of Co-eff. of Correlation.

$x \rightarrow$ $y \downarrow$			80–90	90–100	100–110	110–120	120–130	Total f	fv	fv^2	fuv
	Mid point		85	95	105	115	125				
		$u \rightarrow$ v	−2	−1	0	1	2				
50 – 55	52·5	−3	6 1	9 3	0 7	−15 5	−12 2	18	−54	162	−12
55 – 60	57·5	−1	4 2	4 4	0 10	−7 7	−8 4	27	−27	27	−7
60 – 65	62·5	1	−2 1	−5 5	0 12	10 10	14 7	35	35	35	17
65 – 70	76.5	3	—	−9 3	0 8	18 6	18 3	20	60	180	27
Total f			4	15	37	28	16	100	14	404	25
fu			−8	−15	0	28	32	37			
fu^2			16	15	0	28	64	123			
fuv			8	−1	0	6	12	25			

$$u = \frac{(x-105)}{10}$$

$$v = \frac{(y-60)}{2\cdot5}$$

$$r = \frac{25 - \frac{1}{100}(37)(14)}{\sqrt{123 - \frac{1}{100}(37)^2}\sqrt{404 - \frac{1}{100}(14)^2}} \approx 0{\cdot}09.$$

Ex. 13-7. *Find the variance of the variate*

$u = a_1x_1 + a_2x_2 + ... + a_nx_n$. *(where a's are constants)*

in terms of variances of x_1, x_2 etc.

Sol. Let $\bar{u}, \bar{x}_1, \bar{x}_2$, etc., be the expected values of the variates.

Then $\bar{u} = E(u) = E(a_1x_1 + a_2x_2 + ... + a_nx_n)$

i.e., $\bar{u} = a_1\bar{x}_1 + a_2\bar{x}_2 + ... + a_n\bar{x}_n$

$$\begin{aligned}
\therefore \quad \text{Var}(u) &= E\{u-u\}^2 = E\{a_1(x_1-\bar{x}_1) + a_2(x_2-\bar{x}_2) + ... + a_n(x_n-\bar{x}_n)\}^2 \\
&= E\{a_1^2(x_1-\bar{x}_1)^2 + a_2^2(x_2-\bar{x}_2)^2 + ... + a_n^2(x_n-\bar{x}_n)^2 \\
&\quad + 2a_1a_2(x_1-\bar{x}_1)(x_2-\bar{x}_2) + ...\} \\
&= a_1^2 E(x_1-\bar{x}_1)^2 + a_2^2 E(x_2-\bar{x}_2)^2 + ... \\
&\quad + a_n^2 E(x_n-\bar{x}_n)^2 + 2a_1a_2 E(x_1-\bar{x}_1)(x_2-\bar{x}_2) + ... \\
&= a_1^2 \text{ var }(x_1) + a_2^2 \text{ var }(x_2) + ... + a_n^2 \text{ var }(x_n) \\
&\quad + 2a_1a_2 \text{ Cov }(x_1, x_2) + ...
\end{aligned}$$

Ex. 13-8. *If* σ_x^2, σ_y^2 *and* σ_{x-y}^2 *be the variances of x, y and x – y respectively, show that*

$$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}.$$

Sol.
$$\begin{aligned}
\sigma_{x-y}^2 = \text{var}(x-y) &= \text{var}(x) + \text{var}(y) - 2\text{ cov}(x, y) \\
&= \sigma_x^2 + \sigma_y^2 - 2r_{xy}\sigma_x\sigma_y \\
\therefore \quad r_{xy} &= \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}.
\end{aligned}$$

Ex. 13-9. *Find the correlation co-efficient between x and a – x.*

Sol. Let $u = a - x$

Then $\bar{u} = a - \bar{x}$

where $\bar{u}, \bar{x}$ are expected values

$$\begin{aligned}
\therefore \quad \text{var}(u) &= E(u-\bar{u})^2 = E(x-\bar{x})^2 = \sigma^2 \text{ (say)} \\
\text{cov}(x, u) &= E\{(x-\bar{x})(u-\bar{u})\} = -E(x-\bar{x})^2 = -\sigma^2 \\
\therefore \quad r_{xy} &= \frac{-\sigma^2}{\sigma,\sigma} = -1.
\end{aligned}$$

Ex. 13-10. *Find the correlation co-efficient between x + y and x – y (it is given that x and y are uncorrelated).*

Sol. Let $u = x + y, \; v = x - y$

Then $\bar{u} = \bar{x} + \bar{y}, \; \bar{v} = \bar{x} - \bar{y}$

where $\bar{u}, \bar{v}$ etc. are A. Ms.

$$\begin{aligned}
\therefore \quad \text{cov}(u, v) &= E\{(u-\bar{u})(v-\bar{v})\} \\
&= E[\{(x-\bar{x})+(y-\bar{y})\}\{(x-\bar{x})-(y-\bar{y})\}]
\end{aligned}$$

where σ_x, σ_y etc., are s.ds.

Also var(u) =

i.e., σ_u^2 =

and σ_v^2 =

$\therefore$ If r be the correlation c

r =

Ex. 13-11. *If x and y are two c*

co-efficient r, show that the correl

Sol. Let $u = x + y$ and σ the

Then $\bar{u}$ =

where $\bar{u}, \bar{x}, \bar{y}$ are A.Ms.

Now var(u) =

cov (u, x) =

$\therefore$ Correlation co-efficien

r_{ux} =

Ex. 13-12. *If* x_1, x_2 *and* x_3 *be*

the correlation co-efficient betwe

Sol. Let $\bar{u}$ and $\bar{v}$ be the A.I

Then $\bar{u}$ =

where $\bar{x}_1, \bar{x}_2$ *etc.*, are A.Ms. of x

Now var(u) =

where var (x_1) =

Similarly var (v) =

cov (u, v) =

$\therefore$ Correlation co-effcient b

r =

Ex. 13-13. *Two variates x*

correlation. Show that

$U = x \cos\alpha + y \sin\sigma$

have the same variance σ^2 *and z*

$$\frac{(37)(14)}{\sqrt{404 - \frac{1}{100}(14)^2}} \approx 0{\cdot}09.$$

(where a's are constants)

f the variates.

$_n)$

$\bar{x}_n$

$)+a_2(x_2-\bar{x}_2)+\ldots+a_n(x_n-\bar{x}_n)\}^2$

$-\bar{x}_2)^2+\ldots\ldots+a_n^2(x_n-\bar{x}_n)^2$

$)+\ldots\ldots\}$

$-\bar{x}_2)^2+\ldots\ldots$

$_2\,E(x_1-\bar{x}_1)(x_2-\bar{x}_2)+\ldots\ldots$

$+\ldots\ldots+a_n^2\,\mathrm{var}(x_n)$

s of x, y and x − y respectively, show

$2\,\mathrm{cov}(x,y)$

σ_y

een x and a − x.

$\bar{x})^2 = \sigma^2$ (say)

$= -E(x-\bar{x})^2 = -\sigma^2$

ween x + y and x − y (it is given that x

$\bar{y})\}\{(x-\bar{x})-(y-\bar{y})\}]$

$$= E\{(x-\bar{x})^2-(y-\bar{y})^2\} = \sigma_x^2 - \sigma_y^2$$

where σ_x, σ_y etc., are s.ds.

Also $\quad \mathrm{var}(u) = \mathrm{var}(x)+\mathrm{var}(y)+2\,\mathrm{cov}(x,y)$

i.e., $\quad \sigma_u^2 = \sigma_x^2+\sigma_y^2$

and $\quad \sigma_v^2 = \sigma_x^2+\sigma_y^2$

∴ If r be the correlation co-efficient between u and v,

$$r = \frac{\sigma_x^2-\sigma_y^2}{\sigma_x^2+\sigma_y^2}.$$

Ex. 13-11. *If x and y are two correlated variables with the same s.d. and the correlation co-efficient r, show that the correlation co-efficient between x and x + y is* $\sqrt{\dfrac{1+r}{2}}$.

Sol. Let $u = x+y$ and σ the s.d. of x or y.

Then $\quad \bar{u} = \bar{x}+\bar{v}$

where $\quad \bar{u}, \bar{x}, \bar{y}$ are A.Ms.

Now
$$\begin{aligned}
\mathrm{var}(u) &= \mathrm{var}(x)+\mathrm{var}(y)+2\,\mathrm{cov}(x,y) = 2\sigma^2+2r\sigma^2\\
&= 2\sigma^2(1+r)\\
\mathrm{cov}(u,x) &= E\{(u-\bar{u})(x-\bar{x})] = E[\{(x-\bar{x})+(y-\bar{y})\}(x-\bar{x})]\\
&= E(x-\bar{x})^2+E\{(x-\bar{x})(y-\bar{y})\} = \sigma^2+\mathrm{cov}(x,y)\\
&= \sigma^2(1+r)
\end{aligned}$$

∴ Correlation co-efficient between u and x is given by

$$r_{ux} = \frac{\sigma^2(1+r)}{\sigma^2\sqrt{2(1+r)}} = \sqrt{\frac{1+r}{2}}.$$

Ex. 13-12. *If x_1, x_2 and x_3 be uncorrelated variables each having the same s.d., obtain the correlation co-efficient between $u = x_1+x_2$ and $v = x_2+x_3$.*

Sol. Let $\bar{u}$ and $\bar{v}$ be the A.Ms. of u and v respectively.

Then $\quad \bar{u} = \bar{x}_1+\bar{x}_2$ and $\bar{v} = \bar{x}_2+\bar{x}_3$

where $\bar{x}_1$, $\bar{x}_2$ *etc.*, are A.Ms. of x_1 and x_2 *etc.*, respectively.

Now $\quad \mathrm{var}(u) = \mathrm{var}(x_1)+\mathrm{var}(x_2) = 2\sigma^2$

where $\quad \mathrm{var}(x_1) = \sigma^2$ etc.

Similarly $\quad \mathrm{var}(v) = 2\sigma^2$

$$\begin{aligned}
\mathrm{cov}(u,v) &= E\{(u-\bar{u})(v-\bar{v})\} = E[\{(x_1-\bar{x}_1)+(x_2-\bar{x}_2)\}\{(x_2-\bar{x}_2)+(x_3-\bar{x}_3)\}]\\
&= E(x_2-\bar{x}_2)^2 \qquad (\because\ \mathrm{cov}(x_1, x_3)=0 \text{ etc., as } x_1, x_2, x_3 \text{ are uncorrelated}).\\
&= \sigma^2
\end{aligned}$$

∴ Correlation co-effcient between u and v is given by

$$r = \frac{\sigma^2}{2\sigma^2} = \frac{1}{2}.$$

Ex. 13-13. *Two variates x and y have zero means, the same variance σ^2 and zero correlation. Show that*

$$U = x\cos\alpha + y\sin\sigma \text{ and } V = x\sin\alpha - y\cos\alpha$$

have the same variance σ^2 and zero correlation.

Sol. Since x and y have zero correlation, cov $(x, y) = 0$ *i.e.,* $E(xy) = 0$.

Now
$$\text{var}(U) = \cos^2\alpha\,\text{var}(x) + \sin^2\alpha\,\text{var}(y) = \sigma^2$$
$$\text{var}(V) = \sin^2\alpha\,\text{var}(x) + \cos^2\alpha\,\text{var}(y) = \sigma^2$$
$$\text{cov}(U, V) = E[(U-\overline{U})(V-\overline{V})]$$

Now
$$\overline{U} = \cos\alpha\,\bar{x} + \sin\alpha\,\bar{y} = 0$$

and
$$\overline{V} = \bar{x}\sin\alpha - \bar{y}\cos\alpha = 0$$

$\therefore$
$$\begin{aligned}\text{cov}(U, V) &= E(UV) = E\{(x\cos\alpha + y\sin\alpha)(x\sin\alpha - y\cos\alpha)\}\\ &= \cos\alpha\sin\alpha\,E(x^2) - \sin\alpha\cos\alpha\,E(y^2)\\ &= \cos\alpha\sin\alpha\,\sigma^2 - \sin\alpha\cos\alpha\,\sigma^2 = 0.\end{aligned}$$

Ex. 13-14. *If $u = ax + by$; $v = cx + dy$, show that*

$$\begin{vmatrix}\text{var}(u) & \text{cov}(u,v)\\ \text{cov}(u,v) & \text{var}(v)\end{vmatrix} = \begin{vmatrix}a & b\\ c & d\end{vmatrix}^2 \begin{vmatrix}\text{var}(x) & \text{cov}(x,y)\\ \text{cov}(x,y) & \text{var}(y)\end{vmatrix}$$

Sol. $u = ax + by$ and $v = cx + dy$

$\Rightarrow$ $\bar{u} = a\bar{x} + b\bar{y}$ and $\bar{v} = c\bar{x} + d\bar{y}$

$\therefore$ $u - \bar{u} = a(x-\bar{x}) + b(y-\bar{y})$ and $v - \bar{v} = c(x-\bar{x}) + d(y-\bar{y})$

$\therefore$
$$\begin{aligned}\sigma_u^2 &= \text{var}(u) = E\{a(x-\bar{x}) + b(y-\bar{y})\}^2\\ &= a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\,\text{cov}(x, y)\end{aligned}$$
$$\begin{aligned}\sigma_v^2 &= \text{var}(v)\\ &= c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x, y)\end{aligned}$$
$$\begin{aligned}\text{cov}(u, v) &= E(u-\bar{u})(v-\bar{v})\\ &= E\{a(x-\bar{x}) + b(y-\bar{y})\}\{c(x-\bar{x}) + d(y-\bar{y})\}\\ &= ac\,\sigma_x^2 + bd\,\sigma_y^2 + (ad + bc)\,\text{cov}(x, y)\end{aligned}$$

$\therefore$
$$\Delta_{uv} = \begin{vmatrix}\sigma_u^2 & \text{cov}(u,v)\\ \text{cov}(u,v) & \sigma_y^2\end{vmatrix}$$

$$= \begin{vmatrix}\sigma^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\,\text{cov}(x,y) & ac\sigma_x^2 + bd\sigma_y^2 + (ad+bc)\,\text{cov}(x,y)\\ ac\sigma_x^2 + bd\sigma_y^2 + (ad+bc)\,\text{cov}(x,y) & c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x,y)\end{vmatrix}$$

Apply
$$R_1 \to R_1 - \frac{a}{c}R_2$$

$$= \begin{vmatrix}\dfrac{bc-ad}{c}(b\sigma_y^2 + a\,\text{cov}(x,y)\} & \dfrac{bc-ad}{c}\{d\sigma_y^2 + c.\text{cov}(x,y)\}\\ ac\sigma_x^2 + bd\sigma_y^2 + (ad+bc)\,\text{cov}(x,y) & c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x,y)\end{vmatrix}$$

$$= \frac{bc-ad}{c}\begin{vmatrix}b\sigma_y^2 + a\,\text{cov}(x,y) & d\sigma_y^2 + c.\text{cov}(x,y)\\ ac\sigma_x^2 + bd\sigma_y^2 + (ad+bc)\,\text{cov}(x,y) & c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x,y)\end{vmatrix}$$

Apply
$$R_1 \to R_2 - d.\,R_1$$

$$= \frac{bc-ad}{c}\begin{vmatrix}b\sigma_y^2 + a\,\text{cov}(x,y) & d\sigma_y^2 + c.\text{cov}(x,y)\\ c\{a\sigma_x^2 + b\,\text{cov}(x,y)\} & c\{c\sigma_x^2 + d\,\text{cov}(x,y)\}\end{vmatrix}$$

$= 0$ *i.e.*, $E(xy) = 0$.

$\alpha\ \text{var}(y) = \sigma^2$

$\alpha\ \text{var}(y) = \sigma^2$

$+ y \sin\alpha)(x \sin\alpha - y\cos\alpha)\}$

$\sin\alpha\cos\alpha\ E(y^2)$

$\alpha\cos\alpha\ \sigma^2 = 0.$

$\text{cov}(x,y)$

$\text{var}(y)$

$-\bar{x}) + d(y-\bar{y})$

$)+d(y-\bar{y})\}$

$\text{v}(x, y)$

$ac\sigma_x^2 + bd\sigma_y^2 + (ad+bc)\text{cov}(x,y)$

$c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x,y)$

$\frac{bc-ad}{c}\{d\sigma_y^2 + c.\text{cov}(x,y)\}$

$\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x,y)$

$d\sigma_y^2 + c.\text{cov}(x,y)$

$c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\,\text{cov}(x,y)$

$+ c.\text{cov}(x,y)$

$+ d\,\text{cov}(x,y)\}$

$$= (bc-ad)\begin{vmatrix} b\sigma_y^2 + a\,\text{cov}(x,y) & d\sigma_y^2 + c.\text{cov}(x,y) \\ a\sigma_x^2 + b\,\text{cov}(x,y) & c\sigma_x^2 + d\,\text{cov}(x,y)\end{vmatrix}$$

Apply $\qquad C_1 \to C_1 - \dfrac{b}{d}C_2$

$$= (bc-ad)\begin{vmatrix} \dfrac{(ad-bc)}{d}\text{cov}(x,y) & d\sigma_y^2 + c.\text{cov}(x,y) \\ \dfrac{(ad-bc)}{d}\sigma_x^2 & c\sigma_x^2 + d.\text{cov}(x,y)\end{vmatrix}$$

$$= -\frac{(bc-ad)^2}{d}\begin{vmatrix} \text{cov}(x,y) & d\sigma_y^2 + c.\text{cov}(x,y) \\ \sigma_x^2 & c\sigma_x^2 + d.\text{cov}(x,y)\end{vmatrix}$$

Apply $\qquad C_2 \to C_2 - c\,C_1$

$$= -(bc-ad)^2\begin{vmatrix} \text{cov}(x,y) & \sigma_y^2 \\ \sigma_x^2 & \text{cov}(x,y)\end{vmatrix}$$

$$= (bc-ad)^2\begin{vmatrix} \sigma_x^2 & \text{cov}(x,y) \\ \text{cov}(x,y) & \sigma_y^2\end{vmatrix}$$

$$= (bc-ad)^2\,\Delta_{xy}.$$

Ex. 13-15. *If $u = ax + by$ and $v = bx - ay$, where x and y represent deviations from the respective means and if the co-efficient of correlation between x and y is r and u and v are uncorrelated, show that*

$$\sigma_u \,.\, \sigma_v = (a^2 + b^2)\,\sigma_x\,\sigma_y\,\sqrt{1-r^2}$$

where σ_u, σ_v etc., are s.d. of u. v. etc.

Sol. Now

$$\begin{aligned}
\sigma_u^2 &= \text{var}(u) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\,\text{cov}(x,y) \\
\sigma_v^2 &= \text{var}(v) = b^2\sigma_x^2 + a^2\sigma_y^2 - 2ab\,\text{cov}(x,y) \\
\text{cov}(u,v) &= E\{(ax+by)(bx-ay)\} \\
&= ab(\sigma_x^2 - \sigma_y^2) + (b^2 - a^2)\,\text{cov}(x,y)
\end{aligned}$$

$$\begin{aligned}
\therefore \quad \sigma_u^2\sigma_v^2 - \text{cov}^2(u,v) &= (a^2+b^2)^2\{\sigma_x^2\sigma_y^2 - \text{cov}^2(x,y)\} \\
&= (a^2+b^2)^2\sigma_x^2\sigma_y^2\left\{1 - \frac{\text{cov}^2(x,y)}{\sigma_x^2\sigma_y^2}\right\} \\
&= (a^2+b^2)^2\sigma_x^2\sigma_y^2(1-r^2)
\end{aligned}$$

But $\qquad \text{cov}(u,v) = 0$

$$\therefore \quad \sigma_u\sigma_v = (a^2+b^2)\,\sigma_x\sigma_y\sqrt{1-r^2}\,.$$

Ex. 13-16. *x_1, x_2 are two variates with variances σ_1^2 and σ_2^2 respectively and ρ is the correlation co-efficient between them. Determine the values of the constants a and b which are independent of ρ such that $x_1 + ax_2$ and $x_1 + bx_2$ are uncorrelated.*

Sol. Let $\qquad u = x_1 + ax_2$ and $v = x_1 + bx_2$.

$$\begin{aligned}
\therefore \quad \bar{u} &= \bar{x}_1 + a\bar{x}_2 \text{ and } \bar{v} = \bar{x}_1 + b\bar{x}_2 \\
\therefore \quad \text{cov}(u,v) &= E\{(u-\bar{u})(v-\bar{v})\} \\
&= E\{(x_1-\bar{x}_1) + a(x_2-\bar{x}_2)\}\{(x_1-\bar{x}_1) + b(x_2-\bar{x}_2)\} \\
&= E(x_1-\bar{x}_1)^2 + (a+b)E\{(x_1-\bar{x}_1)(x_2-\bar{x}_2)\} + abE(x_2-\bar{x}_2)^2 \\
&= \sigma_1^2 + (a+b)\,\text{cov}(x_1,x_2) + ab\sigma_2^2
\end{aligned}$$

$$= \sigma_1^2 + ab\sigma_2^2 + (a+b)\,\rho\sigma_1\sigma_2$$

Now $\quad \text{cov}(u, v) = 0$

$\therefore \quad (\sigma_1^2 + ab\sigma_2^2) + (a+b)\,\rho\sigma_1\sigma_2 = 0$

Since a and b are independent of ρ, a and b are given by

$$\sigma_1^2 + ab\sigma_2^2 = 0$$

and

$$a + b = 0$$

$$\therefore \quad a = -b = \frac{\sigma_1}{\sigma_2}.$$

Ex. 13-17. *If x and y are two variates each with mean zero and variance unity and $r_{xy} = r\ (\neq -1)$, find 'b' so that 'x + y' and 'x + by' may be uncorrelated.*

Sol. Let $\quad u = x + y \quad$ and $\quad v = x + by$

Then $\quad \bar{u} = \bar{x} + \bar{y} = 0 \quad$ and $\quad \bar{v} = \bar{x} + b\bar{y} = 0$

$$\begin{aligned}
\therefore \quad 0 = \text{cov}(u, v) &= E(x+y)(x+by) \\
&= E(x^2) + (1+b)\,E(xy) + bE(y^2) \\
&= (1+b)(1+r)
\end{aligned}$$

$$\therefore \quad 1 + b = 0 \quad \text{as} \quad r \neq -1.$$

$$\Rightarrow \quad b = -1.$$

Ex. 13-18. *If x and y are independent random variates, show that*

$$r(x+y,\, x-y) = r^2(x,\, x+y) - r^2(y,\, x+y)$$

where $r(x+y, x-y)$ denotes the co-efficient of correlation between $x + y$ and $x - y$.

Sol. Since x and y are independent,

$$\text{cov}(x, y) = 0 \qquad \ldots(1)$$

Put $\quad X = x + y, \quad Y = x - y$

$$\begin{aligned}
\therefore \quad \text{var}(X) &= \text{var}(x) + \text{var}(y) + 2\,\text{cov}(x, y) \\
&= \sigma_x^2 + \sigma_y^2
\end{aligned}$$

and

$$\begin{aligned}
\text{var}(Y) &= \text{var}(x) + \text{var}(y) - 2\,\text{cov}(x, y) \\
&= \sigma_x^2 + \sigma_y^2
\end{aligned}$$

Now $\quad \bar{X} = \bar{x} + \bar{y}, \quad \bar{Y} = \bar{x} - \bar{y}$

$$\begin{aligned}
\therefore \quad \text{cov}(X, Y) &= E\{X - \bar{X}\}\{Y - \bar{Y}\} \\
&= E\{(x-\bar{x}) + (y-\bar{y})\}\{(x-\bar{x}) - (y-\bar{y})\} \\
&= E\{(x-\bar{x})^2 - (y-\bar{y})^2\} \\
&= \sigma_x^2 - \sigma_y^2 \\
r(X, Y) &= \frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2} \\
\text{cov}(x, X) &= E(x-\bar{x})(X-\bar{X}) \\
&= E(x-\bar{x})\{(x-\bar{x}) + (y-\bar{y})\} \\
&= E(x-\bar{x})^2 + E(x-\bar{x}) - (y-\bar{y}) \\
&= \sigma_x^2 + \text{cov}(x, y) \\
&= \sigma_x^2 \\
\text{cov}(y, X) &= E(y-\bar{y})(X-\bar{X}) \\
&= E(y-\bar{y})\{(x-\bar{x}) + (y-\bar{y})\} \\
&= E(x-\bar{x})(y-\bar{y}) + E(y-\bar{y})^2 \\
&= \sigma_y^2
\end{aligned}$$

$$\therefore \quad r(x, X) = \frac{\sigma_x^2}{\sqrt{\sigma_x^2 + \sigma_y^2,\sigma_x}} = \frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

and $\quad r(y, X) = \dfrac{\ldots}{\sqrt{\sigma_x^2 \ldots}}$

$\therefore \quad r(X, Y) = r^2(x, \ldots$

Ex. 13-19. *x_1, x_2, x_3 are three varia... efficient between any two of them is r. If*

$$\bar{x} = \frac{x_1 + \ldots}{\ldots}$$

show that $\quad \text{var}(\bar{x}) = \dfrac{\sigma^2}{3}\ (\ldots$

Deduce that $\quad r \geq -\dfrac{1}{2}.$

Sol. $\quad \bar{x} = \dfrac{x_1 + \ldots}{\ldots}$

$E(\bar{x}) = \dfrac{\bar{x}_1 + \ldots}{\ldots}$

$\therefore \quad \text{var}(\bar{x}) = E\{\bar{x} - \ldots$

$= E\left\{ \ldots \right.$

$= \dfrac{1}{9} E\{ \ldots$

$+\ 2(x \ldots$

$= \dfrac{1}{9}\ \{3 \ldots$

$= \dfrac{1}{9}\ \{3 \ldots$

$= \dfrac{1}{3}\ \sigma^2 \ldots$

Since $\quad \text{var}(\bar{x}) \geq 0,$

$1 + 2r \geq 0$

$\Rightarrow \quad r \geq -\dfrac{1}{2}.$

Ex. 13.20. *x and y are independent r... 1. Find 'a' so that the correlation co-effi...*

Sol. By given

$\bar{x} = 0 = \bar{y} \ldots$

$\text{cov}(x, y) = 0$

Put $\quad X = x + ay \ldots$

Then $\quad \bar{X} = 0 = \bar{Y}$

$\therefore \quad \text{Cov}(X, Y) = E(XY) \ldots$

$= E\{(x - \ldots$

$= E(x^2) \ldots$

$= 1 + a \ldots$

$\text{var}(X) = \text{var}(x) \ldots$

and $$r(y, X) = \frac{\sigma_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

$$\therefore \quad r(X, Y) = r^2(x, X) - r^2(y, X).$$

Ex. 13-19. *x_1, x_2, x_3 are three variables each with variance σ^2 and the correlation co-efficient between any two of them is r. If*

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3},$$

show that $$\text{var}(\bar{x}) = \frac{\sigma^2}{3}(1 + 2r)$$

Deduce that $$r \geq -\frac{1}{2}.$$

Sol. $$\bar{x} = \frac{x_1 + x_2 + x_3}{3}$$

$$E(\bar{x}) = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3}$$

$$\begin{aligned}
\therefore \quad \text{var}(\bar{x}) &= E\{\bar{x} - E(\bar{x})\}^2 \\
&= E\left\{\frac{(x_1 - \bar{x}_1) + (x_2 - \bar{x}_2) + (x_3 - \bar{x}_3)}{3}\right\}^2 \\
&= \frac{1}{9} E\{(x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2 + (x_3 - \bar{x}_3)^2 \\
&\quad + 2(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + 2(x_2 - \bar{x}_2)(x_3 - \bar{x}_3) \\
&\quad + 2(x_1 - \bar{x}_1)(x_3 - \bar{x}_3)\} \\
&= \frac{1}{9}\{3\sigma^2 + 2\,\text{cov}(x_1, x_2) + 2\,\text{cov}(x_2, x_3) + 2\,\text{cov}(x_1, x_3)\} \\
&= \frac{1}{9}\{3\sigma^2 + 6\sigma^2 r\} \quad (\because \ \text{cov}(x_1, x_2) = \sigma^2 r \text{ etc.}) \\
&= \frac{1}{3}\sigma^2(1 + 2r)
\end{aligned}$$

Since $$\text{var}(\bar{x}) \geq 0,$$
$$1 + 2r \geq 0$$
$$\Rightarrow \quad r \geq -\frac{1}{2}.$$

Ex. 13.20. *x and y are independent random variables each with mean zero and variance 1. Find 'a' so that the correlation co-efficient between x + ay* and *x + y is maximum.*

Sol. By given

$$\bar{x} = 0 = \bar{y}, \ \sigma_x = \sigma_y = 1$$
$$\text{cov}(x, y) = 0$$

Put $$X = x + ay, \ Y = x + y$$

Then $$\bar{X} = 0 = \bar{Y}$$

$$\begin{aligned}
\therefore \quad \text{Cov}(X, Y) &= E(XY) \\
&= E\{(x + ay)(x + y)\} \\
&= E(x^2) + (1 + a)E(xy) + aE(y^2) \\
&= 1 + a
\end{aligned}$$

$$\text{var}(X) = \text{var}(x) + a^2\,\text{var}(y)$$

$$= 1 + a^2$$

$$\text{var}(Y) = \text{var}(x) + \text{var}(y) = 2$$

$$\therefore \quad r_{XY} = \frac{1+a}{\sqrt{2}\sqrt{1+a^2}}$$

Now maximum value of $r_{XY} = 1$

$$\therefore \quad \frac{1+a}{\sqrt{2(1+a^2)}} = 1$$

$$\Rightarrow \quad 1 + a^2 + 2a = 2 + 2a^2$$

i.e., $\quad a^2 - 2a + 1 = 0$

$$\Rightarrow \quad a = 1.$$

Ex. 13-21. *If x and y are uncorrelated random variates with means zero and variances σ_1^2 and σ_2^2 respectively. Show that the correlation co-efficient between*

$$u = x \sin\alpha + y\cos\alpha$$

and

$$v = x\cos\alpha - y\sin\alpha$$

is

$$\frac{\sigma_1^2 - \sigma_2^2}{\left\{\left(\sigma_1^2 - \sigma_2^2\right)^2 + 4\sigma_1^2\sigma_2^2\,\text{cosec}^2\,2\alpha\right\}^{1/2}}.$$

Sol. By given

$$\bar{x} = 0 = \bar{y},$$

$$\text{var}(x) = \sigma_1^2,\ \text{var}(y) = \sigma_2^2$$

and $\quad \text{cov}(x, y) = 0$

Now

$$u = x\sin\alpha + y\cos\alpha$$

$$v = x\cos\alpha - y\sin\alpha$$

$$\therefore \quad \bar{u} = 0 = \bar{v}$$

$$\begin{aligned}
\sigma_u^2 &= \text{var}(u) = \sin^2\alpha\ \text{var}(x) + \cos^2\alpha\ \text{var}(y)\\
&= \sin^2\alpha\,\sigma_1^2 + \cos^2\alpha\,\sigma_2^2\\
\sigma_v^2 &= \text{var}(v) = \cos^2\alpha\ \text{var}(x) + \sin^2\alpha\ \text{var}(y)\\
&= \cos^2\alpha\,\sigma_1^2 + \sin^2\alpha\,\sigma_2^2
\end{aligned}$$

$$\begin{aligned}
\therefore \quad \sigma_u^2\sigma_v^2 &= (\sin^2\alpha\sigma_1^2 + \cos^2\alpha\sigma_2^2)(\cos^2\alpha\sigma_1^2 + \sin^2\alpha\sigma_2^2)\\
&= \sin^2\alpha\cos^2\alpha\{\sigma_1^4 + \sigma_2^4 - 2\sigma_1^2\sigma_2^2\}\\
&\quad + \sigma_1^2\sigma_2^2\{\sin^4\alpha + \cos^4\alpha + 2\sin^2\alpha\cos^2\alpha\}\\
&= \sin^2\alpha\cos^2\alpha(\sigma_1^2 - \sigma_2^2)^2 + \sigma_1^2\sigma_2^2
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(u, v) &= E(uv)\\
&= E\{(x\sin\alpha + y\cos\alpha)(x\cos\alpha - y\sin\alpha)\}\\
&= E\{(x^2 - y^2)\sin\alpha\cos\alpha + xy(\cos^2\alpha - \sin^2\alpha)\}\\
&= (\sigma_1^2 - \sigma_2^2)\sin\alpha\cos\alpha + \text{cov}(x, y)\{\cos^2\alpha - \sin^2\alpha\}\\
&= (\sigma_1^2 - \sigma_2^2)\sin\alpha\cos\alpha
\end{aligned}$$

$$r_{uv} = \frac{\left(\sigma_1^2 - \sigma_2^2\right)\cos\alpha\sin\alpha}{\left\{\sin^2\alpha\cos^2\alpha\left(\sigma_1^2 - \sigma_2^2\right)^2 + \sigma_1^2\sigma_2^2\right\}^{1/2}}$$

$$= \frac{\left(\sigma_1^2 - \sigma_2^2\right)}{\left\{\left(\sigma_1^2 - \sigma_2^2\right)^2 + 4\sigma_1^2\sigma_2^2\,\text{cosec}^2\,2\alpha\right\}^{1/2}}.$$

Ex. 13-22. *If x and y are random*
show that

$$u = x\sin\alpha +$$

$$v = y\sin\alpha -$$

are uncorrelated if

$$\tan 2\alpha = \frac{2\gamma\sigma_x\sigma_y}{\sigma_y^2 - \sigma_x^2}$$

Sol. By given

$$u = x\sin\alpha +$$

$$v = y\sin\alpha -$$

$$\bar{u} = \bar{x}\sin\alpha +$$

$$\bar{v} = \bar{y}\sin\alpha -$$

$$\text{cov}(u, v) = E\{(u - \bar{u}$$

$$= E\{(x - \bar{x}$$

$$= E(-\cos$$

$$+ (x - \bar{x})$$

$$= \sin\alpha\cos$$

$$= \sin\alpha\cos$$

Now u and v are uncorrelated if

$$\text{cov}(u, v) = 0$$

$$\Rightarrow \quad \sin\alpha\cos\alpha(\sigma_y^2 - \sigma_x^2) - \gamma\sigma$$

$$\Rightarrow \quad \tan 2\alpha = \frac{2\gamma\sigma_x\sigma}{\sigma_y^2 - \sigma}$$

Ex. 13-23. *Let x be a binomial va*

(*i*) $E\left(\frac{x}{n} - p\right)^2 = \frac{pq}{n}$

(*ii*) $Cov\left(\frac{x}{n}, \frac{n-x}{n}\right) = -\frac{pq}{n}$, q

Sol. (*i*) Since x is B.V. with para

$$\bar{x} = np \quad \text{ar}$$

$$\therefore \quad E\left(\frac{x}{n} - p\right)^2 = E\left(\frac{x -}{}\right.$$

$$= \frac{1}{n^2}E($$

$$= \frac{1}{n^2}\text{va}$$

$$= \frac{npq}{n^2} =$$

(*ii*) Let $\quad u = \frac{x}{n}, \quad v$

Then $\quad \bar{u} = \frac{\bar{x}}{n}, \quad \bar{v} =$

Ex. 13-22. *If x and y are random variates with correlation co-efficient γ between them, show that*

$$u = x \sin \alpha + y \cos \alpha$$
$$v = y \sin \alpha - x \cos \alpha$$

are uncorrelated if

$$\tan 2\alpha = \frac{2\gamma\sigma_x \sigma_y}{\sigma_y^2 - \sigma_x^2}.$$

Sol. By given

$$\begin{aligned}
u &= x \sin \alpha + y \cos \alpha \\
v &= y \sin \alpha - x \cos \alpha \\
\bar{u} &= \bar{x} \sin \alpha + \bar{y} \cos \alpha \\
\bar{v} &= \bar{y} \sin \alpha - \bar{x} \cos \alpha \\
\text{cov}(u, v) &= E\{(u - \bar{u})(v - \bar{v})\} \\
&= E\{(x - \bar{x})\sin\alpha + (y - \bar{y})\cos\alpha\}\ \{(y - \bar{y})\sin\alpha - (x - \bar{x})\cos\alpha\} \\
&= E(-\cos\alpha \sin\alpha(x - \bar{x})^2 + \sin\alpha\cos\alpha\,(y - \bar{y})^2 \\
&\quad + (x - \bar{x})(y - \bar{y})(\sin^2\alpha - \cos^2\alpha)\} \\
&= \sin\alpha\cos\alpha(\sigma_y^2 - \sigma_x^2) - \text{cov}\,(x, y)\cos 2\alpha \\
&= \sin\alpha\cos\alpha\,(\sigma_y^2 - \sigma_x^2) - \gamma\sigma_x\sigma_y \cos 2\alpha
\end{aligned}$$

Now u and v are uncorrelated if

$$\text{cov}\,(u, v) = 0$$

$$\Rightarrow \quad \sin\alpha\cos\alpha\,(\sigma_y^2 - \sigma_x^2) - \gamma\sigma_x\sigma_y\cos 2\alpha = 0$$

$$\Rightarrow \quad \tan 2\alpha = \frac{2\gamma\sigma_x\sigma_y}{\sigma_y^2 - \sigma_x^2}.$$

Ex. 13-23. *Let x be a binomial variate with parameters n and p. Show that*

(*i*) $E\left(\frac{x}{n} - p\right)^2 = \frac{pq}{n}$

(*ii*) $Cov\left(\frac{x}{n}, \frac{n-x}{n}\right) = -\frac{pq}{n},\ q = 1 - p.$

Sol. (*i*) Since x is B.V. with parameters n, p

$$\bar{x} = np \quad \text{and} \quad \text{var}\,(x) = npq$$

$$\begin{aligned}
\therefore \quad E\left(\frac{x}{n} - p\right)^2 &= E\left(\frac{x - np}{n}\right)^2 \\
&= \frac{1}{n^2} E(x - np)^2 \\
&= \frac{1}{n^2}\,\text{var}\,(x) \\
&= \frac{npq}{n^2} = \frac{pq}{n}
\end{aligned}$$

(*ii*) Let $\quad u = \frac{x}{n}, \quad v = \frac{n-x}{n}$

Then $\quad \bar{u} = \frac{\bar{x}}{n}, \quad \bar{v} = \frac{n - \bar{x}}{n}$

$$\therefore \quad \operatorname{cov}\left(\frac{x}{n}, \frac{n-x}{n}\right) = \operatorname{cov}(u, v)$$

$$= E\{(u-\bar{u})(v-\bar{v})\}$$

$$= E\left\{\left(\frac{x}{n}-\frac{\bar{x}}{n}\right)\left(\frac{n-x}{n}-\frac{n-\bar{x}}{n}\right)\right\}$$

$$= -E\left(\frac{x-\bar{x}}{n}\right)^2 = -\frac{1}{n^2}E(x-\bar{x})^2$$

$$= -\frac{\operatorname{var}(x)}{n^2} = -\frac{pq}{n}.$$

13.3. Rank Correlation

13.3-1. Non-Repeated Ranks

Let n be the number of individuals which are ranked according to two different characters A and B. Let x and y be the ranks w.r.t. A and B respectively. Assuming that ranks are not repeated in either series, both x and y take the same values 1, 2, ... *n*.

Then $$\Sigma x = \Sigma y = 1 + 2 + \ldots + n = \frac{n(n+1)}{2}$$

and $$\Sigma x^2 = \Sigma y^2 = 1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\therefore \quad \operatorname{var}(x) = \operatorname{var}(y) = \frac{1}{n}\Sigma x^2 - \left(\frac{\Sigma x}{n}\right)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \left\{\frac{n(n+1)}{2n}\right\}^2 = \frac{n^2-1}{12}$$

Let $$d = x \sim y$$

$$\therefore \quad \Sigma d^2 = \Sigma(x-y)^2 = \Sigma\{(x-\bar{x})-(y-\bar{y})\}^2 \qquad (\because \bar{x} = \bar{y})$$

where $\bar{x}$ and $\bar{y}$ are A.Ms.

$$\therefore \quad \frac{1}{n}\Sigma d^2 = \frac{1}{n}\Sigma(x-\bar{x})^2 + \frac{1}{n}\Sigma(y-\bar{y})^2 - 2\frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})$$

$$= \operatorname{var}(x) + \operatorname{var}(y) - 2\operatorname{cov}(x, y)$$

$$= \frac{n^2-1}{12} + \frac{n^2-1}{12} - 2\operatorname{cov}(x, y)$$

$$\therefore \quad \operatorname{cov}(x, y) = \frac{n^2-1}{12} - \frac{1}{2n}\Sigma d^2.$$

∴ Correlation co-efficient between x and y is given by

$$r = \frac{\operatorname{cov}(x,y)}{(\text{s.d. of } x)\,(\text{s.d. of } y)} = 1 - \frac{6}{n(n^2-1)}\Sigma d^2$$

'r' is called Spearman's rank correlation co-efficient.

Ex. 13-24. *Calculate Spearman's rank correlation co-efficient from the following data. Two numbers within brackets denote the ranks of the students in papers A and B respectively.*

(1, 1); (2, 10); (3, 3); (4, 4); (5, 5); (6, 7); (7, 2); (8, 6); (9, 8); (10, 11); (11, 15); (12, 9); (13, 14); (14, 12); (15, 16); (16, 13).

Sol. Let R_1 and R_2 be the ranks

Calculation

R_1	R_2	$d = R_1 \sim R_2$
1	1	0
2	10	8
3	3	0
4	4	0
5	5	0
6	7	1
7	2	5
8	6	2

∴ Spearman's Rank Correlatio

$r = 1 -$

$= 0{\cdot}8.$

Ex. 13-25. *Ten competitors in*

data :

First Judge	:	1	6
Second Judge	:	3	5
Third Judge	:	6	4

Use the method of rank correl

approach to common likings in voic

Sol. Let R_1, R_2, and R_3 be the r

Calculation

Comp. No.	R_1	R_2	R_3
1	1	3	6
2	6	5	4
3	5	8	9
4	10	4	8
5	3	7	1
6	2	10	2
7	4	2	3
8	9	1	10
9	7	6	5
10	8	9	7

∴ Rank co-eff. of correlation b

$= 1 -$

Rank Co-eff. of correlation betv

$= 1 -$

$\ldots - \frac{n-\bar{x}}{n}\Big)\Big\}$

$\frac{1}{n^2} E(x-\bar{x})^2$

ked according to two different characters spectively. Assuming that ranks are not e values 1, 2, ... *n.*

$= \frac{n(n+1)}{2}$

$+ n^2 = \frac{n(n+1)(2n+1)}{6}$

$\left(\frac{\Sigma x}{n}\right)^2$

$\left\{\frac{n(n+1)}{2n}\right\}^2 = \frac{n^2-1}{12}$

$\bar{x})-(y-\bar{y})\}^2 \qquad (\because \bar{x}=\bar{y})$

$(y-\bar{y})^2 - 2\frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})$

2 cov (x, y)

2 cov (x, y)

given by

$= 1 - \frac{6}{n(n^2-1)}\Sigma d^2$

ent.

ition co-efficient from the following data e students in papers A and B respectively

); (8, 6); (9, 8); (10, 11); (11, 15); (12, 9

Sol. Let R_1 and R_2 be the ranks for A and B respectively.

Calculation of Co-eff. of Rank Correlation

R_1	R_2	$d = R_1 \sim R_2$	d^2	R_1	R_2	$R_1 \sim R_2 = d$	d^2
1	1	0	0	9	8	1	1
2	10	8	64	10	11	1	1
3	3	0	0	11	15	4	16
4	4	0	0	12	9	3	9
5	5	0	0	13	14	1	1
6	7	1	1	14	12	2	4
7	2	5	25	15	16	1	1
8	6	2	4	16	13	3	9
							136

∴ Spearman's Rank Correlation Co-efficient is given by

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6.(136)}{16.255}$$

$$= 0{\cdot}8.$$

Ex. 13-25. *Ten competitors in a voice test are ranked by three judges in the following data :*

First Judge	:	1	6	5	10	3	2	4	9	7	8
Second Judge	:	3	5	8	4	7	10	2	1	6	9
Third Judge	:	6	4	9	8	1	2	3	10	5	7

Use the method of rank correlation to gauge which pair of judges have the nearest approach to common likings in voice.

Sol. Let R_1, R_2, and R_3 be the ranks due to three judges respectively.

Calculation of Rank Co-eff. of Correlation

Comp. No.	R_1	R_2	R_3	$R_1 - R_2 = d_{12}$	d_{12}^2	$R_2 - R_3 = d_{23}$	d_{23}^2	$R_1 - R_3 = d_{13}$	d_{13}^2
1	1	3	6	−2	4	−3	9	−5	25
2	6	5	4	1	1	1	1	2	4
3	5	8	9	−3	9	−1	1	−4	16
4	10	4	8	6	36	−4	16	2	4
5	3	7	1	−4	16	6	36	2	4
6	2	10	2	−8	64	8	64	0	0
7	4	2	3	2	4	−1	1	1	1
8	9	1	10	8	64	−9	81	−1	1
9	7	6	5	1	1	1	1	2	4
10	8	9	7	−1	1	2	4	1	1
					200		214		60

∴ Rank co-eff. of correlation between first and second judge

$$= 1 - \frac{1200}{10.99} \approx -0{\cdot}212$$

Rank Co-eff. of correlation between first and third judge

$$= 1 - \frac{(6)(60)}{10.99} \approx 0{\cdot}636$$

and rank co-efficient of correlation between second and third judge

$$= 1 - \frac{6.214}{10.99} \approx -0{\cdot}297.$$

Since the correlation between first and second judges is –ve, opinions regarding voice test are opposite of each other. Similarly the opinions of second and third judge are opposite of each other. But the opinions of first judge and third are of similar type as their correlation is positive *i.e.,* their likings and dislikings are very much common.

Hence first and third judges have nearest approach to the common likings.

13.3-2. Repeated Ranks

In this case two or more individuals are bracketed equal in either or both classifications. Here, common ranks are given to the bracketed individuals. This common ranks is the average of the ranks which these individuals would have assumed had they been slightly different in ranks from each other.

The ranks co-eff. of correlation when $t_1, t_2 \ldots. t_p$ figures are given same ranks, is given by

$$r = 1 - \frac{6(\Sigma d^2 + T)}{n(n^2 - 1)}$$

where $$T = \sum_{i=1}^{p} \frac{1}{12}\left(t_i^3 - t_i\right).$$

Ex. 13-26. *Find spearman's rank correlation co-efficient for the data given below :*

Students	:	1	2	3	4	5	6	7	8	9	10	11	12
Marks in Exam. A	:	15	13	17	14	18	12	20	16	18	17	19	21
Marks in Exam. B	:	18	16	18	15	19	16	18	15	21	17	18	20

Sol.

Calculation of Rank Co.eff. Correlation

S.N.	*Ranks in A* R_1	*Ranks in B* R_2	$R_1 \sim R_2$ d	d^2
1	9	5·5	3·5	12·25
2	11	9·5	1·5	2·25
3	6·5	5·5	1·0	1·00
4	10	11·5	1·5	2·25
5	4·5	3	1·5	2·25
6	12	9·5	2·5	6·25
7	2	5·5	3·5	12·25
8	8	11·5	3·5	12·25
9	4·5	1	3·5	12·25
10	6·5	8	1·5	2·25
11	3	5·5	2·5	6·25
12	1	2	1	1·00
				72·50

Here in paper A, two students have got 18 marks each and two 17 marks each. While marking ranks, ranks 1, 2, 3 are given to students getting marks 21, 20, 19, ranks 4th and 5th are to be given to students getting 18 each. As they have got equal marks, their ranks should be same and hence each is given the rank.

$$\frac{4+5}{2} = 4{\cdot}5.$$

l third judge

ges is –ve, opinions regarding voice
second and third judge are opposite
re of similar type as their correlation
ch common.
n to the common likings.

qual in either or both classifications.
viduals. This common ranks is the
ave assumed had they been slightly

res are given same ranks, is given by

efficient for the data given below :

7	8	9	10	11	12
20	16	18	17	19	21
18	15	21	17	18	20

Correlation

$R_1 \sim R_2$ d	d^2
3·5	12·25
1·5	2·25
1·0	1·00
1·5	2·25
1·5	2·25
2·5	6·25
3·5	12·25
3·5	12·25
3·5	12·25
1·5	2·25
2·5	6·25
1	1·00
	72·50

s each and two 17 marks each. While
ng marks 21, 20, 19, ranks 4th and 5th
ve got equal marks, their ranks should

Similarly the ranks of students getting 17 marks each is

$$\frac{6+7}{2} = 6{\cdot}5 \text{ each}$$

In paper B, there are four students getting marks 18 each and the ranks to be given to them are 4, 5, 6, 7.

$\therefore$ Rank of each student getting marks 18

$$= \frac{4+5+6+7}{4} = \frac{22}{4} = 5{\cdot}5.$$

Similarly ranks of each student getting 16 marks in B

$$= \frac{9+10}{2} = 9{\cdot}5$$

and rank of each student getting 15 marks

$$= \frac{11+12}{2} = 11{\cdot}5$$

As in a paper A, two students get the same rank 4·5 and two students get the same rank 6·5, for paper A we have $t_1 = 2,\ t_2 = 2$.

In paper B, there are 4 students getting rank 5·5, two students getting rank 9·5 and two students getting rank 11·5.

$$\therefore \quad t_3 = 4,\ t_4 = 2,\ t_5 = 2.$$

$$\therefore \quad T = \frac{1}{12}\{2^3-2\} + \frac{1}{12}\{2^3-2\} + \frac{1}{12}\{4^3-4\} + \frac{1}{12}\{2^3-2\} + \frac{1}{12}\{2^3-2\}$$

$$= \frac{1}{3}\{8-2\} + \frac{1}{12}\{64-4\}$$

$$= 2+5 = 7$$

$$\therefore \quad r = 1 - \frac{6(72{\cdot}5+7)}{(12)(143)}$$

$$= 1 - \frac{79{\cdot}5}{286} = \frac{206{\cdot}5}{286} = 0{\cdot}722.$$

Ex. 13-27. *The co-efficients of rank correlation of the marks obtained by 10 students in Physics and Maths. was found to be 0·4. It was later on discovered that the difference in ranks for one student was wrongly taken as 2 instead of 3. Find the correct co-efficient of rank correlation.*

Sol. Here $r = 0{\cdot}4,\ n = 10$

$$\therefore \quad 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 0{\cdot}4$$

$$\Rightarrow \quad 0{\cdot}6 = \frac{6\Sigma d^2}{10(99)}$$

$$\Rightarrow \quad \Sigma d^2 = 99.$$

Now, corrected value of $\Sigma d^2 = 99 - 4 + 9 = 104$.

Corrected value of rank correlation co-efficient

$$= 1 - \frac{6(104)}{10 \cdot 99} = 0{\cdot}37.$$

13.3.3. Limits for the Ranks Correlation Co-efficient

The formula for rank correlation co-efficient is

$$r = 1 - \frac{6.\Sigma d^2}{n(n^2-1)}$$

where r is the rank correlation co-efficient. Now, r is maximum when Σd^2 is minimum, which is so only when each d is minimum *i.e.,* zero. This is achieved only when ranks of each individual are same in either classification.

$\therefore$ Minimum value of $\Sigma d^2 = 0$

$\therefore$ Maximum value of $r = 1$

r is minimum, when Σd^2 is maximum. This is achieved only when the ranks in two classifications are in reverse order *i.e.,* if the rank of an individuals in one classification is r, its rank in other classification is $n - \overline{r-1} = n + 1 - r$. In this case corresponding value of d is

$|n - (2r - 1)|$

$\therefore$ Maximum value of Σd^2

$$= \sum_{r=1}^{n} \{n - (2r - 1)\}^2$$

$$= \sum_{r=1}^{n} \{n^2 + (4r^2 - 4r + 1) - 2n(2r - 1)\}$$

$$= \sum_{r=1}^{n} \{(n + 1)^2 - 4(n + 1) r + 4r^2\}$$

$$= n(n + 1)^2 - 4(n + 1) \sum_{r=1}^{n} r + 4 \sum_{r=1}^{n} r^2$$

$$= n(n + 1)^2 - 4(n + 1) \cdot \frac{n(n+1)}{2} + 4\frac{n(n+1)(2n+1)}{6}$$

$$= - n(n + 1)^2 + \frac{2}{3} n(n + 1)(2n + 1)$$

$$= \frac{1}{3} n(n + 1) \{4n + 2 - 3(n + 1)\}$$

$$= \frac{1}{3} n(n^2 - 1)$$

Min. value of $r = 1 - \dfrac{6}{n(n^2-1)} \cdot \dfrac{1}{3} n(n^2 - 1)$

$= -1.$

13.4. Regression and Lines of Regression

In case there is some relationship between the variates, the points of the scater diagram will be more or less concentrated round a curve. This curve is called curve of regression. From this curve, it is possible to estimate one of the variables (the dependent variable) from the other (the independent variable). This process of estimation is often referred to as, regression. If y (or x) is estimated from x (or y), regression curve is of y on x (or x on y).

In case this curve is a straight line, it is called the line of regression and the regression is said to be linear.

Evidently the line of regression is the straight line which gives the 'best fit in the least square sense' to the given data.

In case y is treated as dependent and x independent variable, the line of regression is called the 'line of regression of y on x' and gives the best estimate of y for any given value of

x. In the contrary case it is called
of x for any given value of y.

13.4-1. Equations of Lines of I

Consider the bivariate freq.

$x \rightarrow$

$y \rightarrow$

$f \rightarrow$

where $f_1 + f_2 + ... f_n = N$.

The line of regression is the
distribution.

Let $y = mx + c$ be the equ
unknowns to be determined by t

Let $Y_i = r$

and $S =$

According to the method o
minimum.

The normal equations are :

$0 =$

i.e., $m \sum_{i=1}^{n} f_i x_i + c \sum_{i=1}^{n} f_i$

i.e., $m\bar{x} + c = \bar{y}$

and $0 =$

i.e., $m \sum_{i=1}^{n} f_i x_i^2 + c \sum_{i=1}^{n} f_i$

Now by def,

$\sigma_x^2 =$

$=$

$\sum_{i=1}^{n} f_i x_i^2 =$

and $\mu_{11} =$

$=$

$=$

t

maximum when Σd^2 is minimum,
iis is achieved only when ranks of

when the ranks in two classifications
e classification is r, its rank in other
nding value of d is

$- 2n\,(2r - 1)\}$

$r + 4r^2\}$

$r + 4\sum_{r=1}^{n} r^2$

$\frac{n+1)}{2} + 4\frac{n(n+1)(2n+1)}{6}$

$)\,(2n + 1)$

$(n + 1)\}$

$1)$

tes, the points of the scater diagram
curve is called curve of regression.
ables (the dependent variable) from
estimation is often referred to as,
ion curve is of y on x (or x on y).
ine of regression and the regression

which gives the 'best fit in the least

nt variable, the line of regression is
t estimate of y for any given value of

x. In the contrary case it is called the 'line of regression of x on y' and gives the best estimate of x for any given value of y.

13.4-1. Equations of Lines of Regression

Consider the bivariate freq. dist.

$$\begin{array}{l} x \rightarrow \\ y \rightarrow \\ f \rightarrow \end{array} \begin{pmatrix} x_1 & x_2 & \cdots\cdots & x_n \\ y_1 & y_2 & \cdots\cdots & y_n \\ f_1 & f_2 & \cdots\cdots & f_n \end{pmatrix}$$

where $f_1 + f_2 + \ldots.. f_n = N$.

The line of regression is the straight line best fitted in the least square sense to the given distribution.

Let $y = mx + c$ be the equation of line of regression of y on x where m and c are unknowns to be determined by the method of least squares.

Let $$Y_i = mx_i + c$$

and $$S = \sum_{i=1}^{n} f_i (Y_i - y_1)^2 = \sum_{i=1}^{n} f_i (mx_i + c - y_i)^2$$

According to the method of least squares, m and c are to be determined so that S is minimum.

The normal equations are :

$$0 = \frac{\partial S}{\partial c} = \sum_{i=1}^{n} 2f_i (mx_i + c - y_i)$$

i.e., $$m\sum_{i=1}^{n} f_i x_i + c\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} f_i y_i$$

i.e., $$m\bar{x} + c = \bar{y} \qquad \ldots(1)$$

and $$0 = \frac{\partial S}{\partial m} = \sum_{i=1}^{n} 2f_i (mx_i + c - y_i)(x_i)$$

i.e., $$m\sum_{i=1}^{n} f_i x_i^2 + c\sum_{i=1}^{n} f_i x_i = \sum_{i=1}^{n} f_i x_i y_i \qquad \ldots(2)$$

Now by def,

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i^2 + \bar{x}^2 - 2x_i\,\bar{x}\}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \bar{x}^2$$

$$\sum_{i=1}^{n} f_i x_i^2 = n\left(\sigma_x^2 + \bar{x}^2\right)$$

and $$\mu_{11} = \text{cov}\,(x, y) = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})\,(y_i - \bar{y})$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \{x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}.\bar{y}\}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i y_i - \bar{x}.\bar{y}$$

$$\therefore \qquad \sum_{i=1}^{n} f_i\, x_i y_i = N(\mu_{11} + \bar{x}.\bar{y})$$

Substituting in (2)

$$m(\sigma_x^2 + \bar{x}^2) + \bar{x}c = \mu_{11} + \bar{x}\,.\,\bar{y} \qquad \text{...(3)}$$

Solving (1) and (3) for m and c

$$m = \frac{\mu_{11}}{\sigma_x^2} \text{ and } c = \bar{y} - \frac{\mu_{11}}{\sigma_x^2}\bar{x}$$

$\therefore$ Eq. of line of regression of y on x is

$$y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2}(x - \bar{x})$$

Similarly the line of regression of x on y can be shown to have its equation

$$x - \bar{x} = \frac{\mu_{11}}{\sigma_y^2}(y - \bar{y}).$$

Ex. 13-28. *If α is the angle between the two regression lines in the case of two variables x and y show that*

$$\left(\tan\alpha = \frac{1-r^2}{r}\,\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

where the symbols have their usual meanings.

Sol. The equation of two lines of regression are

$$y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2}(x - \bar{x}) \qquad \text{...(1) } (y \text{ on } x)$$

$$x - \bar{x} = \frac{\mu_{11}}{\sigma_y^2}(y - \bar{y}) \qquad \text{...(2) } (x \text{ on } y)$$

Case I. If $\mu_{11} \neq 0$.

$$\text{Slope of (1)} = \frac{\mu_{11}}{\sigma_x^2} = r\frac{\sigma_y}{\sigma_x}$$

$$\text{Slope of (2)} = \frac{\sigma_y^2}{\mu_{11}} = \frac{\sigma_y}{r\sigma_x}$$

where r is the correlation co-efficient between x and y.

$$\text{Now} \qquad \tan\alpha = \frac{\dfrac{1}{r}\dfrac{\sigma_y}{\sigma_x} - \dfrac{r\sigma_y}{\sigma_x}}{1 + \dfrac{\sigma_y^2}{\sigma_x^2}} = \frac{1-r^2}{r}\,\frac{\sigma_y\sigma_x}{\sigma_x^2 + \sigma_y^2}$$

Case II. If $\mu_{11} = 0$.

Lines (1) and (2) become

$$y = \bar{y} \quad \text{and} \quad x = \bar{x}$$

which are parallel to co-ordinate axes

$$\therefore \qquad \alpha = 90°.$$

13.4-2. Regression Co-efficients

The quantities $\dfrac{\mu_{11}}{\sigma_x^2}$ and $\dfrac{\mu_{11}}{\sigma_y^2}$ are called regression co-efficients of 'y on x' and 'x on y' respectively and are denoted by b_{yx} and b_{xy} respectively.

13.4-3. Properties of Regre

(*i*) *The correlation co-eff*
Regression co-efficients

b_{yx} =

$\therefore$ $b_{yx}\,.\,b_{xy}$ =

where r is the correlation co-
(*ii*) *The correlation co-eff*
regression co-efficients.
Regression co-efficients

$b_{yx} = \dfrac{\mu_{11}}{\sigma_x^2}$ =

$\therefore$ $\left|\dfrac{b_{yx}+b_{xy}}{2}\right|$ =

$\therefore$ $\left|\dfrac{b_{yx}+b_{xy}}{2}\right|$ −

=

$\therefore$ $\left|\dfrac{b_{yx}+b_{xy}}{2}\right|$ ≥

Remarks. (1) b_{yx} is the s
the slope of line of regression
(2) b_{yx}, b_{xy} and r are of
Ex. 13-29. *The ages (X*
below :

Age in ye *(X)*
56
42
72
36
63
47
55
49
38
42
68
60

Determine the least squar
co-efficient of Y on X.
Also estimate the blood p

13.4-3. Properties of Regression Co-efficients

(*i*) *The correlation co-efficient is the geometric mean between regression co-efficients.*

Regression co-efficients are given by

$$b_{yx} = \frac{\mu_{11}}{\sigma_x^2} \text{ and } b_{xy} = \frac{\mu_{11}}{\sigma_y^2}$$

$$\therefore \qquad b_{yx} \,.\, b_{xy} = \left(\frac{\mu_{11}}{\sigma_x \sigma_y}\right)^2 = r^2$$

where r is the correlation co-efficient.

(*ii*) *The correlation co-efficient cannot numerically exceed the arithmetic mean between regression co-efficients.*

Regression co-efficients are given by

$$b_{yx} = \frac{\mu_{11}}{\sigma_x^2} = r\frac{\sigma_y}{\sigma_x} \quad \text{and} \quad b_{xy} = \frac{\mu_{11}}{\sigma_y^2} = r\frac{\sigma_x}{\sigma_y}$$

$$\therefore \qquad \left|\frac{b_{yx}+b_{xy}}{2}\right| = |r|\,\frac{\sigma_y^2+\sigma_x^2}{2\sigma_x\sigma_y}$$

$$\therefore \qquad \left|\frac{b_{yx}+b_{xy}}{2}\right| - |r| = |r|\left\{\frac{\sigma_y^2+\sigma_x^2}{2\sigma_x\sigma_y} - 1\right\}$$

$$= |r|\,\frac{(\sigma_y-\sigma_x)^2}{2\sigma_x\sigma_y} \geq 0$$

$$\therefore \qquad \left|\frac{b_{yx}+b_{xy}}{2}\right| \geq |r|$$

Remarks. (1) b_{yx} is the slope of line of regression of y on x and b_{xy} is the reciprocal of the slope of line of regression of x on y.

(2) b_{yx}, b_{xy} and r are of same signs.

Ex. 13-29. *The ages (X) and systolic blood pressures (Y) of 12 women are given below :*

Age in years (*X*)	*Blood Pressure* (*Y*)
56	147
42	125
72	160
36	118
63	149
47	128
55	150
49	145
38	115
42	140
68	152
60	155

Determine the least squares regression line of Y on X and find the value of the regression co-efficient of Y on X.

Also estimate the blood pressure of a woman whose age is 45 years.

Sol.

X	Y	x	y	x^2	xy	
56	147	4	7	16	28	
42	125	– 10	– 15	100	150	
72	160	20	20	400	400	
36	118	– 16	– 22	256	352	
63	149	11	9	121	99	$x = X - 52$
47	128	– 5	– 12	25	60	
55	150	3	10	9	30	$y = Y - 140$
49	145	– 3	5	9	– 15	
38	115	– 14	– 25	196	350	
42	140	– 10	0	100	0	
68	152	16	12	256	192	
60	155	8	15	64	120	
		4	4	1552	1766	

Let the line of regression of y on x be

$$y = a + bx$$

where the co-efficients 'a' and 'b' are given by the equations

$$\Sigma y = na + b\Sigma x$$

and $$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Substituting the values of Σx etc.,

$$4 = 12a + 4b \quad \text{or} \quad 1 = 3a + b$$

and $$1766 = 4a + 1552b \quad \text{or} \quad 883 = 2a + 776b$$

Solving

$$a = -0{\cdot}046 \quad \text{and} \quad b = 1{\cdot}138$$

$\therefore$ The equation of line of regression of y on x is

$$y = (-0{\cdot}046) + (1{\cdot}138)x$$

$\therefore$ The equation of line of regression of Y on X is

$$Y - 140 = -0{\cdot}046 + (1{\cdot}138)(X - 52)$$

or $$Y = (1{\cdot}138)X + (80{\cdot}778)$$

$\therefore$ Regression co-efficient of Y on $X = (1{\cdot}138)$

Now value of x for $X = 45$ is $(45 - 52) = -7$.

$\therefore$ Estimate of y $= -0{\cdot}046 - 7{\cdot}966$
$= -8{\cdot}012$

$\therefore$ Estimate of Y for $X = 45$
$= 140 - 8{\cdot}012$
$= 131{\cdot}988$

Ex. 13-30. *For the following table :*

Age of husbands in years	*Ages of wives in years*					
	10 — 20	20 — 30	30 — 40	40 — 50	50 — 60	Total
15 — 25	6	3	—	—	—	9
25 — 35	3	16	10	—	—	29
35 — 45	—	10	15	7	—	32
45 — 55	—	—	7	10	4	21
55 — 65	—	—	—	4	5	9
Total	9	29	32	21	9	100

2	xy	
6	28	
0	150	
0	400	
6	352	
1	99	$x = X - 52$
5	60	
9	30	$y = Y - 140$
9	– 15	
6	350	
0	0	
6	192	
4	120	
2	1766	

tions

$$1 = 3a + b$$
$$883 = 2a + 776b$$

8

s

2)

ars

40 — 50	50 — 60	Total
—	—	9
—	—	29
7	—	32
10	4	21
4	5	9
21	9	100

Find (*i*) *the co-efficient of correlation,*
(*ii*) *the two regression lines.*

Sol. The calculating table is on page 532

(*i*) $\therefore$ The co-efficient of correlation is given by

$$r = \frac{98 - \frac{1}{100}(-8)(-8)}{\sqrt{122 - \frac{1}{100}(-8)^2}\ \sqrt{122 - \frac{1}{100}(-8)^2}}$$

$$= \frac{9800 - 64}{(12200 - 64)}$$

$$= \frac{9736}{12136} = 0{\cdot}802$$

$\therefore$ (*ii*) $$\bar{u} = \bar{v} = \frac{(-8)}{100} = -0{\cdot}08$$

and $$\sigma_u^2 = \sigma_u^2 = \frac{1}{100}\{122\} - \left(-\frac{8}{100}\right)^2$$

$$= \frac{1}{(100)^2}\{12200 - 64\} = 1{\cdot}2136$$

$\therefore$ $$b_{uv} = b_{vu} = r = 0{\cdot}802.$$

$\therefore$ The equations of lines of regression of v on u and u on v respectively are

$$(v + 0{\cdot}08) = 0{\cdot}802\,(u + 0{\cdot}08)$$

and $$(u + 0{\cdot}08) = 0{\cdot}802\,(v + 0{\cdot}08)$$

$\therefore$ The equation of lines of regression of y on x and x on y respectively are

$$\left(\frac{y-40}{10} + 0\cdot 08\right) = 0{\cdot}802\left(\frac{x-35}{10} + 0\cdot 08\right)$$

or $$y = 0{\cdot}802\,x + 11{\cdot}772$$

and $$\left(\frac{x-35}{10} + 0\cdot 08\right) = (0{\cdot}802)\left(\frac{y-40}{10} + 0\cdot 08\right)$$

or $$x = 0{\cdot}802\,y + 2{\cdot}762.$$

Ex. 13-31. *In a partially destroyed laboratory record of the correlation analysis of data, the following results only are legible : var* $(x) = 9$, *regression lines* $8x - 10y + 66 = 0$ *and* $40x - 18y = 214$.

Supply (*i*) *mean values of x and y.*
(*ii*) *the s.d. of y.*
(*iii*) *the correlation co-efficient between x and y.*

Sol. (*i*) The equations of lines of regression of y on x and x on y respectively are

$$y - \bar{y} = b_{yx}(x - \bar{x}) \quad ...(1)$$

and $$x - \bar{x} = b_{xy}(y - \bar{y}) \quad ...(2)$$

$\therefore$ Thus $\bar{x}$ and $\bar{y}$ are the values of x and y which satisfy both the regression equations simultaneously.

$\therefore$ Solving regression equations

$$\bar{x} = 13 \quad \text{and} \quad \bar{y} = 17$$

(*ii*) It is not given, of the two regression equations which represents the line of regression of y on x. So we assume,

Let x be the variable for the ages of wives and y be the variable for the ages of husbands.

$x \rightarrow$ $y \downarrow$			10–20	20–30	30–40	40–50	50–60	Total f	fv	fv^2	fuv	
	Mid-point		15	25	35	45	55					
		$u \rightarrow$ $v \downarrow$	– 2	– 1	0	1	2					
15 – 25	20	– 2	24 6	6 3	—	—	—	9	–18	36	30	$u = \frac{(x-35)}{10}$
25 – 35	30	–1	6 3	16 16	0 10	—	—	29	–29	29	22	$v = \frac{(y-40)}{10}$
35 –45	40	0	—	0 10	0 15	0 7	—	32	0	0	0	
45 – 55	50	1	—	—	0 7	10 10	8 4	21	21	21	18	
55 – 65	60	2	—	—	—	8 4	20 5	9	18	36	28	
Total f			9	29	32	21	9	100	– 8	122	98	
fu			– 18	– 29	0	21	18	– 8				
fu^2			36	29	0	21	36	122				
fuv			30	22	0	18	28	98				

45 – 55	50	1	—	—	0 7	10 10	8 4	21	21	21	18
55 – 65	60	2	—	—	—	8 4	20 5	9	18	36	28
Total f			9	29	32	21	9	100	–8	122	98
fu			–18	–29	0	21	18	–8			
fu^2			36	29	0	21	36	122			
fuv			30	22	0	18	28	98			

$$8x - 10y + 66 = 0$$

to be the equation representating the line of regression of y on x. Then equation representing the line of regression of x on y must be

$$40x - 18y = 214$$

$\therefore$ Comparing with (1) and (2)

$$b_{yx} = \frac{4}{5} \quad \text{and} \quad b_{xy} = \frac{9}{20}$$

$\therefore$ The co-eff. of correlation r is given by

$$r^2 = b_{yx} \,.\, b_{xy} = \frac{4}{5} \cdot \frac{9}{20} = \frac{9}{25} \ (< 1).$$

Since r^2 comes out to be less than unity, our assumption is correct.
Since b_{yx} and b_{xy} are positive, r must be positive and hence

$$r = \frac{3}{5} = 0{\cdot}6$$

(*iii*) Now $b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$

$$\therefore \quad \frac{4}{5} = \frac{3}{5} \cdot \frac{\sigma_y}{3}$$

or $\quad \sigma_y = 4.$

Ex. 13.32. *Two random variables have the least squares regression line* $3x + 2y - 26 = 0$ *and* $6x + y - 31 = 0$. *Find the mean values and the correlation co-efficient.*

Sol. Solving regression equations

$$\bar{x} = 4 \quad \text{and} \quad \bar{y} = 7$$

Now slopes of regression lines are

$$-\frac{3}{2} \quad \text{and} \quad -6$$

Since $\quad r^2 \le 1,$

$$b_{yx} = -\frac{3}{2} \quad \text{and} \quad b_{xy} = -\frac{1}{6}$$

$$\therefore \quad r^2 = b_{yx} \,.\, b_{xy} = \frac{1}{4} \ (\le 1)$$

$$\therefore \quad r = -0{\cdot}5. \qquad (\because\ b_{yx},\ b_{xy} \text{ are } < 0)$$

Ex. 13.33. *For a bivariate distribution, the lines of regression are* $3x + 12y = 19$ *and* $3y + 9x = 46$. *Find the mean of the distribution and the correlation co-efficient.*

Sol. Solving regression equations

$$\bar{x} = 5 \quad \text{and} \quad \bar{y} = \frac{1}{3}$$

Now slopes of regression lines are $-\dfrac{1}{4}$ and -3

$$\therefore \quad b_{yx} = -\frac{1}{4} \quad \text{and} \quad b_{xy} = -\frac{1}{3}$$

$$\therefore \quad r^2 = b_{yx} \,.\, b_{xy} = \frac{1}{12}$$

$$\therefore \quad r = -\frac{1}{2\sqrt{3}}. \qquad (\because\ b_{yx},\ b_{xy} < 0)$$

Ex. 13-34. *Given that the lines of regression of y on x and x on y are respectively* $y = x$ *and* $4x - y - 3 = 0$ *and the second moment about the origin for x is 2; calculate* (*i*) *the mean*

for x (ii) the mean for y (iii) variance of x (iv) variance of y (v) the correlation co-efficient between x and y.

Sol. Solving regression equations

$$\therefore \quad \bar{x} = 1 = \bar{y}$$

Now slopes of regression lines are 1 and 4

$$\therefore \quad b_{yx} = 1, \quad b_{xy} = \frac{1}{4}$$

$$\therefore \quad r^2 = \frac{1}{4} \text{ i.e., } r = 0{\cdot}5$$

Now for x, $\mu'_2(0) = 2$

$$\therefore \quad \sigma_x^2 = \mu'_2(0) - \bar{x}^2 = 2 - 1 = 1.$$

Also $\quad 1 = b_{yx} = \dfrac{r\sigma_y}{\sigma_x} = \dfrac{1}{2}\sigma_y$

$$\therefore \quad \sigma_y = 2.$$

Ex. 13.35. *Given $x = 4y + 5$ and $y = kx + 4$ are the regression lines of x on y and y on x respectively. Show that $0 < 4k < 1$. If $k = \frac{1}{16}$, find the means of the two variables and the co-efficient of correlation between them.*

Sol. Here $b_{yx} = k$ and $b_{xy} = 4$

$$\therefore \quad r^2 = b_{yx} \,.\, b_{xy} = 4k.$$

Now since r^2 is the square of real quantity, it should be nonegative.

Since $\quad bxy \neq 0, \quad r \neq 0$

also as two lines of regression are different, $r^2 \neq 1$.

$$\therefore \quad 0 < r^2 < 1$$

$$\Rightarrow \quad 0 < 4k < 1$$

When $k = \frac{1}{16}$, the equations of lines of regression become

$$x = 4y + 5 \qquad (x \text{ on } y)$$

$$16y = x + 64 \qquad (y \text{ on } x)$$

Solving regression equations

$$\therefore \quad \bar{x} = 28 \quad \text{and} \quad \bar{y} = 5{\cdot}75$$

Also $\quad r^2 = 4k = 4 \,.\, \dfrac{1}{16} = \dfrac{1}{4}$

$$\therefore \quad r = \frac{1}{2} = 0{\cdot}5. \qquad (b_{yx}, b_{xy} > 0)$$

Ex. 13-36. *The lines of regression obtained in a correlation analysis are*

$$x + 9y = 7 \quad \text{and} \quad y + 4x = 16\frac{1}{3}$$

Find the (*i*) *the co-efficient of correlation*

(*ii*) *the ratio $\sigma_x^2 : \sigma_y^2 : cov(x, y)$.*

Sol. Slopes of regression lines are

$$-\frac{1}{9} \text{ and } -4$$

Since $\quad r^2 \leq 1,$

$$b_{yx} = -\frac{1}{9}, \; b_{xy} = -\frac{1}{4} \qquad \ldots(1)$$

$\therefore \quad r^2 = b_{yx}$ ·

$\therefore \quad r = -\sqrt{}$

$= -\dfrac{1}{6}$

Also $\quad b_{yx} = -\dfrac{1}{9},$

$\therefore$ (1) $\Rightarrow \quad \dfrac{\mu_{11}}{\sigma_x^2} = -\dfrac{1}{9}$

$\Rightarrow \quad \mu_{11} : \sigma$

$-1 : 9$

Ex. 13-37. *For two variables x*

$x + 2y - 5 = 0$

and $\quad 2x + 3y - 8 = 0$

Also $\quad var(x) = 12$

Find $\bar{x}, \bar{y}, \sigma_y$ and r.

Sol. Solving regression equatio

$\bar{x} = 1$ an

Slopes of regression lines are –

$\therefore \quad b_{yx} = -\dfrac{1}{2}$

$\therefore \quad r^2 = \left(-\dfrac{1}{2}\right.$

$\therefore \quad r = -\dfrac{\sqrt{3}}{2}$

Also $b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$ and $\sigma_x = 2$

$\therefore \quad \left(-\dfrac{1}{2}\right) = \left(-\dfrac{\sqrt{3}}{2}\right.$

$\therefore \quad \sigma_y = 2.$

Ex. 13-38. *For 10 observations obtained (in appropriate units) :*

$\Sigma x = 130,$ Σ

Obtain the line of regression of y units.

Sol. Let the equation of the line c

$y = a + bx,$

where the co-efficients 'a' and 'b' are

$\Sigma y = na + b$

and $\quad \Sigma xy = a\Sigma x +$

Substituting the values

$220 = 10a +$

and $\quad 3467 = 130a +$

of y (v) the correlation co-efficient

e regression lines of x on y and y on
: means of the two variables and the

ıld be nonegative.

ı become

(x on y)
(y on x)

$(b_{yx}, b_{xy} > 0)$

correlation analysis are

$\frac{1}{3}$

...(1)

$$\therefore \quad r^2 = b_{yx} . b_{xy} = \frac{1}{36}$$

$$\therefore \quad r = -\sqrt{\frac{1}{36}} \qquad (\because \ b_{yx}, b_{xy} < 0)$$

$$= -\frac{1}{6}$$

Also $\quad b_{yx} = -\frac{1}{9}, \qquad b_{xy} = \frac{\mu_{11}}{\sigma_y^2}$

$$\therefore (1) \Rightarrow \quad \frac{\mu_{11}}{\sigma_x^2} = -\frac{1}{9}, \qquad \frac{\mu_{11}}{\sigma_y^2} = -\frac{1}{4}$$

$$\Rightarrow \quad \mu_{11} : \sigma_x^2 : \sigma_y^2$$
$$-1 : 9 : 4.$$

Ex. 13-37. *For two variables x and y the two regression lines are*

$$x + 2y - 5 = 0$$

and

$$2x + 3y - 8 = 0$$

Also $\quad var(x) = 12$

Find $\bar{x}, \bar{y}, \sigma_y$ *and r.*

Sol. Solving regression equations

$$\bar{x} = 1 \text{ and } \bar{y} = 2.$$

Slopes of regression lines are $-\frac{1}{2}$ and $-\frac{2}{3}$

$$\therefore \quad b_{yx} = -\frac{1}{2} \text{ and } b_{xy} = -\frac{3}{2}$$

$$\therefore \quad r^2 = \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right) = \frac{3}{4} \ (< 1)$$

$$\therefore \quad r = -\frac{\sqrt{3}}{2} \qquad (\because \ b_{yx}, b_{xy} < 0)$$

Also $b_{yx} = r\frac{\sigma_y}{\sigma_x}$ and $\sigma_x = 2\sqrt{3}$

$$\therefore \quad \left(-\frac{1}{2}\right) = \left(-\frac{\sqrt{3}}{2}\right)\frac{\sigma_y}{2\sqrt{3}}$$

$$\therefore \quad \sigma_y = 2.$$

Ex. 13-38. *For 10 observations on price (x) and supply (y) the following data were obtained (in appropriate units) :*

$$\Sigma x = 130, \ \Sigma y = 220, \ \Sigma x^2 = 2288, \ \Sigma y^2 = 5506, \ \Sigma xy = 3467.$$

Obtain the line of regression of y on x and estimate the supply when the price is 16 units.

Sol. Let the equation of the line of regression of y on x be

$$y = a + bx,$$

where the co-efficients 'a' and 'b' are given by the normal equations

$$\Sigma y = na + b\Sigma x$$

and

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Substituting the values

$$220 = 10a + 130b \text{ or } 22 = a + 13b$$

and

$$3467 = 130a + 2288b$$

$\therefore$ $a = 8{\cdot}8$ and $b = 1{\cdot}015$

$\therefore$ The equation of line of regression of y on x is

$$y = 8{\cdot}8 + 1{\cdot}015x$$

$\therefore$ Estimate of supply (y) when the price (x) is 16 units

$$= 8{\cdot}8 + 16{\cdot}240 = 25{\cdot}04.$$

Ex. 13-39. *From the data given below estimate the most likely height of a father whose son's height is 70″.*

Fathers : Mean height is 67″ with a s.d. of 3·5″
Sons : Mean height is 65″ with a s.d. of 2·5″

Co-efficient of correlation between the heights of fathers and sons is + 0·8.

Sol. Let y be the variable corresponding to the height of fathers and x be the variable for the son's heights.

Then $\bar{x} = 65$, $\bar{y} = 67$, $\sigma_x = 2{\cdot}5$, $\sigma_y = 3{\cdot}5$, and $r_{xy} = 0{\cdot}8$.

$$\therefore \quad b_{yx} = \frac{(0{\cdot}8)(3{\cdot}5)}{(2{\cdot}5)} = 1{\cdot}12$$

$\therefore$ The equation of line of regression of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

or $$y - 67 = 1{\cdot}12(x - 65)$$

or $$y = 1{\cdot}12x - 5{\cdot}8$$

$\therefore$ Most likely height of a father whose son's height is 70″

$$= \text{Estimate of } y \text{ for } x = 70$$
$$= 78{\cdot}4 - 5{\cdot}8 = 72{\cdot}6.$$

Ex. 13-40. *The following statistical co-efficients were deduced in the course of an examination of the relationship between yield of wheat and the amount of rainfall.*

	Yield in lbs (per acre)	*Annual Rainfall (in inches)*
Mean	985·0	12·8
s.d.	70·1	1·6

r(between yield and rainfall) = + 0·52.

From the above data, calculate (i) the most likely yield of wheat per acre when the annual rainfall is 9·2″ and (ii) the probable annual rainfall for yield of 1,400 *lbs. per acre.*

Sol. Let y be the variable for yield and x be the variable for annual rainfall.

Then $\bar{x} = 12{\cdot}8$, $\bar{y} = 985{\cdot}20$, $\sigma_x = 1{\cdot}6$, $\sigma_y = 70{\cdot}1$, and $r_{xy} = 0{\cdot}52$

$$\therefore \quad b_{yx} = \frac{(0{\cdot}52)(70{\cdot}1)}{(1{\cdot}6)} = 22{\cdot}7825$$

and $$b_{xy} = \frac{(0{\cdot}52)(1{\cdot}6)}{70{\cdot}1} = 0{\cdot}01187$$

$\therefore$ The equations of lines of regression are

$$y - 985 = 22{\cdot}7825(x - 12{\cdot}8) \qquad (y \text{ on } x)$$

and $$x - 12{\cdot}8 = (0{\cdot}01187)(y - 985) \qquad (x \text{ on } y)$$

$\therefore$ The most likely yield of wheat per acre when the annual rainfall is 9·2″.

$$= 985 + (22{\cdot}7825)(-3{\cdot}6) = 902{\cdot}983$$
$$= 903$$

and the probable annual rainfall for yield of 1,400 lbs. per acre

$$= 12{\cdot}8 + (0{\cdot}01187)(415)$$
$$= 17{\cdot}72605 \simeq 17{\cdot}7''.$$

.s

16 units

? *most likely height of a father whose*

with a s.d. of 3·5″
with a s.d. of 2·5″
'fathers and sons is + 0·8.
eight of fathers and x be the variable

$_{xy} = 0{\cdot}8$.

x is

height is 70″
70

nts were deduced in the course of an
eat and the amount of rainfall.

Annual Rainfall
(in inches)
12·8
1·6

ikely yield of wheat per acre when the
rainfall for yield of 1,400 *lbs. per acre.*
e variable for annual rainfall.
70·1, and $r_{xy} = 0{\cdot}52$

22·7825

·01187

(y on x)
(x on y)
5)
hen the annual rainfall is 9·2″.
– 3·6) = 902·983

) lbs. per acre
(415)

Ex. 13-41. *The following data give the correlation co-efficient, means and s.d. of rainfall and yield of paddy in a certain tract :*

	Yield (per acre) in lbs	*Annual Rainfall (in inches)*
Mean	973·5	18·3
s.d.	38·4	2·0

Co-efficient of correlation $= 0{\cdot}58$

Estimate the most likely yield of paddy when the annual rainfall is 22″, other factors being assumed to remain the same.

Sol. Let y be the variable for yield and x be the variable for annual rainfall. Then

$$\bar{x} = 18{\cdot}3,\ \bar{y} = 973{\cdot}5,\ \sigma_x = 2{\cdot}0,\ \sigma_y = 38{\cdot}4 \quad \text{and} \quad r_{xy} = 0{\cdot}58.$$

$$\therefore \qquad b_{yx} = \frac{(0\cdot 58)\,(38\cdot 4)}{(2\cdot 0)} = 11{\cdot}136$$

$\therefore$ The equation of line of regression of y on x is

$$y - 973{\cdot}5 = 11{\cdot}136\,(x - 18{\cdot}3)$$

$\therefore$ Estimate of the most likely yield of paddy when the annual rainfall is 22″

$$= \text{Estimate of } y \text{ for } x = 22$$
$$= 973{\cdot}5 + (11{\cdot}136)\,(3{\cdot}7)$$
$$= 973{\cdot}5 + 41{\cdot}2032 = 1014{\cdot}7032 \simeq 1014{\cdot}7.$$

Ex. 13-42. *If a number x is chosen at random from among the integers* 1, 2, 3, 4 *and number y is chosen from among these at least as large as x, prove that*

$$cov\,(x, y) = 5/8$$

Also find the line of regression of x on y.

Sol. Since x is to be selected at random from the integers

$$1, 2, 3, 4$$

prob. of x taking each of these values is $\frac{1}{4}$.

Now, when x takes value 1, y is to be chosen out of 1, 2, 3, 4.

$\therefore$ Conditional prob. of y taking each of these values is $\frac{1}{4}$.

$\therefore$ Prob. of each of the pairs

$$(1, 1), (1, 2), (1, 3)\ (1, 4) \text{ is}$$

$$\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$$

Now, when x takes value 2, y is to be chosen out of 2, 3, 4.

$\therefore$ Conditional prob. of y taking value $1 = 0$,

and conditional prob. of y taking each of the values 2, 3, $4 = \frac{1}{3}$

$\therefore$ Prob. of the pairs

$$(2, 1), (2, 2), (2, 3), (2, 4) \text{ are}$$

$$0,\ \frac{1}{4}\cdot\frac{1}{3} = \frac{1}{12}, \frac{1}{12}, \frac{1}{12}$$

when x takes values 3, y is to be chosen out of 3, 4.

$\therefore$ Conditional prob. of y taking each of the values 1, 2 is zero and conditional prob. of y taking each of the values 3, 4 is $\frac{1}{2}$.

$\therefore$ Prob. of the pairs

(3, 1), (3, 2), (3, 3), (3, 4)

are $$0, 0, \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}, \frac{1}{8}.$$

When x takes value 4, y can take only one value 4.

$\therefore$ Conditional prob. of y taking each of values 1, 2, 3 is zero and the conditional prob. of y taking value 4 is 1.

$\therefore$ Prob. of the pairs

(4, 1), (4, 2), (4, 3), (4, 4)

are $$0, 0, 0, \frac{1}{4} \cdot 1 = \frac{1}{4}$$

Thus, the bivariate distribution is

$y \rightarrow$ / $x \downarrow$	1	2	3	4
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
2	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
3	0	0	$\frac{1}{8}$	$\frac{1}{8}$
4	0	0	0	$\frac{1}{4}$

The calculating table is on next page

(*i*) $$\text{cov}(x, y) = \Sigma pxy - (\Sigma px)(\Sigma py)$$
$$= \frac{420}{48} - \frac{10}{4} \cdot \frac{156}{48}$$
$$= \frac{5}{8}$$

(*ii*) Let eq. of line of regression of x on y is
$$x = a + by$$

Normal equations are
$$\Sigma px = a + b\Sigma py$$
and $$\Sigma pxy = a\Sigma py + b\Sigma py^2$$

Substituting values, equations reduce to
$$\frac{10}{4} = a + b\left(\frac{156}{48}\right)$$
and $$\frac{420}{48} = \frac{156}{48}a + b \cdot \frac{548}{48}$$

$y \rightarrow$ / $x \downarrow$	1	2
1	$\boxed{1/16}$ 1/16	$\boxed{2/16}$ 1/16
2	0	$\boxed{4/12}$ 1/12
3	0	0
4	0	0
p	1/16	7/48
py	1/16	14/48
py^2	1/16	28/48
pxy	1/16	22/48

i.e., $120 = 48$

and $420 = 15$

$\therefore$ $a = 0{\cdot}1$

$\therefore$ Eq. of line of regression of

$x = 0{\cdot}1$

13.4-4. Standard Errors of Estin

Find the standard errors of es

Sol. The eq. of line of regress

$y - \bar{y} = \frac{\mu_1}{\sigma_x^2}$

Let (x_i, y_i), $i = 1, 2 \ldots n$ be the

Let $Y_i = \bar{y}$ -

The standard error of estimate

$S_y^2 = \frac{1}{N} \sum_{i=1}^{n} f_i \{Y_i -$

where $N = \sum_{i=1}^{n} f_i$

$= \frac{1}{N} \sum_{i=1}^{n} f_i \left\{ \frac{\mu_1}{\sigma_x^2} \right.$

$= \frac{1}{N} \sum_{i=1}^{n} f_i \left\{ \left(\frac{\mu}{\sigma} \right. \right.$

ues 1, 2 is zero

4 is $\frac{1}{2}$.

ł.

es 1, 2, 3 is zero and the conditional

3	4
$\frac{1}{16}$	$\frac{1}{16}$
$\frac{1}{12}$	$\frac{1}{12}$
$\frac{1}{8}$	$\frac{1}{8}$
0	$\frac{1}{4}$

Cal. of Covariance

$y \rightarrow$ / $x \downarrow$	1	2	3	4	p	px	px^2	pxy
1	$\boxed{1/16}$ 1/16	$\boxed{2/16}$ 1/16	$\boxed{3/16}$ 1/16	$\boxed{4/16}$ 1/16	1/4	1/4	1/4	10/16
2	0	$\boxed{4/12}$ 1/12	$\boxed{6/12}$ 1/12	$\boxed{8/12}$ 1/12	1/4	2/4	4/4	18/12
3	0	0	$\boxed{9/8}$ 1/8	$\boxed{12/8}$ 1/8	1/4	3/4	9/4	21/8
4	0	0	0	$\boxed{16/4}$ 1/4	1/4	4/4	16/4	16/4
p	1/16	7/48	13/48	25/48	1	10/4	30/4	420/48
py	1/16	14/48	39/48	100/48	156/48			
py^2	1/16	28/48	117/48	400/48	548/48			
pxy	1/16	22/48	87/48	308/48	420/48			

i.e., $120 = 48a + 156b$

and $420 = 156a + 548b$

$\therefore$ $a = 0{\cdot}13, \quad b = 0{\cdot}73$

$\therefore$ Eq. of line of regression of x on y is

$$x = 0{\cdot}13 + 0{\cdot}73y.$$

13.4-4. Standard Errors of Estimate

Find the standard errors of estimate of y and x respectively.

Sol. The eq. of line of regression of y on x is

$$y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2}(x - \bar{x})$$

Let (x_i, y_i), $i = 1, 2 \ldots n$ be the variate value pair occurring with frequency f_i.

Let $$Y_i = \bar{y} + \frac{\mu_{11}}{\sigma_x^2}(x_i - \bar{x})$$

The standard error of estimate of y is given by

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{n} f_i \{Y_i - y_i\}^2$$

where $$N = \sum_{i=1}^{n} f_i$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \left\{\frac{\mu_{11}}{\sigma_x^2}(x_i - \bar{x}) - (y_i - \bar{y})\right\}^2$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \left\{\left(\frac{\mu_{11}}{\sigma_x^2}\right)^2 (x_i - \bar{x})^2 + (y_i - \bar{y})^2 - 2\left(\frac{\mu_{11}}{\sigma_x^2}\right)(x_i - \bar{x})(y_i - \bar{y})\right\}$$

$$= \frac{\mu_{11}^2}{\sigma_x^2} + \sigma_y^2 - 2\,\frac{\mu_{11}^2}{\sigma_x^2}$$

$$= \sigma_y^2 \left\{1 - \left(\frac{\mu_{11}}{\sigma_x \sigma_y}\right)^2\right\} = \sigma_y^2\,(1 - r^2)$$

$$\therefore \qquad S_y = \sigma_y\,(1 - r^2)^{1/2}$$

Similarly standard error of estimate of x is given by

$$S_x = \sigma_x\,(1 - r^2)^{1/2}$$

Note. If $r = \pm 1, S_x = S_y = 0$

$\therefore$ All points lie on both lines of regression and hence two regression lines coincide and thus there is a linear functional relation between the variates x and y.

As $r^2 \to 1, S_x^2 \to 0$ and $S_y^2 \to 0$, *i.e.*, as r^2 comes nearer to unity, the points are closer to lines of regression which are nearer to coincidence.

$\therefore$ The departure of r^2 from unity can be taken as a measure of departure of the relationship between the two variates from linearity.

Ex. 13-43. *For a given bivariate dist. find the straight line for which the sum of the squares of the normal deviations is minimum.*

Sol. Consider the bivariate dist.

$$\begin{matrix} x \to \\ y \to \\ f \to \end{matrix} \begin{pmatrix} x_1\, x_2 \,\ldots\ldots\ldots\, x_n \\ y_1\, y_2 \,\ldots\ldots\ldots\, y_n \\ f_1\, f_2 \,\ldots\ldots\ldots\, f_n \end{pmatrix}$$

and let the equation of the straight line be

$$x \cos\alpha + y \sin\alpha - p = 0 \qquad \ldots(1)$$

The normal deviation of an observed value pair (x_i, y_i) from the line is the length of perpendicular from the point (x_i, y_i) upon the line *i.e.*, $x_i \cos\alpha + y_i \sin\alpha - p$.

Let $$S = \sum_{i=1}^{n} f_i\,(x_i \cos\alpha + y_i \sin\alpha - p)^2$$

Normal eqs. are

$$0 = \frac{\partial S}{\partial \alpha} = \sum_{i=1}^{n} 2f_i\,(x_i \cos\alpha + y_i \sin\alpha - p)\,(-x_i \sin\alpha + y_i \cos\alpha) \qquad \ldots(2)$$

and $$0 = \frac{\partial S}{\partial p} = \sum_{i=1}^{n} -2f_i(x_i \cos\alpha + y_i \sin\alpha - p) \qquad \ldots(3)$$

Eqs. (3) and (2) are equivalent to eqs.

$$\bar{x} \cos\alpha + \bar{y} \sin\alpha = p \qquad \ldots(4)$$

and $$\cos\alpha \sin\alpha \left\{\sum_{i=1}^{n} f_i y_i^2 - \sum_{i=1}^{n} f_i x_i^2\right\} + \cos 2\alpha \sum_{i=1}^{n} f_i x_i y_i$$

$$+ p \left\{\sin\alpha \sum_{i=1}^{n} f_i x_i - \cos\alpha \sum_{i=1}^{n} f_i y_i\right\} = 0$$

i.e., $$\cos\alpha \sin\alpha \{(\sigma_y^2 + \bar{y}^2) - (\sigma_x^2 + \bar{x}^2)\} + \cos 2\alpha(\mu_{11} + \bar{y}\,\bar{x}) + p\bar{x} \sin\alpha - p\bar{y} \cos\alpha = 0$$

i.e., $$\{\cos\alpha \sin\alpha\,(\sigma_y^2 - \sigma_x^2) + \cos 2\alpha\,\mu_{11}\} + \{\cos\alpha \sin\alpha\,(\bar{y}^2 - \bar{x}^2)$$

$$+ \bar{x}.\bar{y}(\cos^2\alpha - \sin^2\alpha) + p\bar{x} \sin\alpha - p\bar{y} \cos\alpha\} = 0$$

$\sigma_y^2 (1 - r^2)$

e two regression lines coincide and
ates x and y.
earer to unity, the points are closer

as a measure of departure of the

ight line for which the sum of the

...(1)

(x_i, y_i) from the line is the length of
$_i \cos \alpha + y_i \sin \alpha - p$.

$\alpha - p)^2$

$-x_i \sin \alpha + y_i \cos \alpha)$ ...(2)

...(3)

...(4)

$\alpha \sum_{i=1}^{n} f_i x_i y_i$

$\alpha(\mu_{11} + \bar{y}\,\bar{x}) + p\bar{x} \sin\alpha - p\bar{y} \cos \alpha = 0$

$\cos \alpha \sin \alpha\, (\bar{y}^2 - \bar{x}^2)$

$\alpha - \sin^2 \alpha) + p\bar{x} \sin\alpha - p\bar{y} \cos\alpha\} = 0$

i.e.,
$$\left\{\frac{\sin 2\alpha}{2}\left(\sigma_y^2 - \sigma_x^2\right) + \cos 2\alpha \,.\, \mu_{11}\right\} + \{\bar{y} \cos\alpha(\bar{y} \sin\alpha + \bar{x}\cos\alpha - p)$$
$$- \bar{x} \sin\alpha(\bar{x}\cos\alpha + \bar{y}\sin\alpha - p)\} = 0$$

i.e.,
$$\frac{\sin 2\alpha}{2}\left(\sigma_y^2 - \sigma_x^2\right) + \cos 2\,\alpha \,.\, \mu_{11} = 0 \qquad \text{[using (4)]}$$

$$\therefore \qquad \tan 2\,\alpha = \frac{2\mu_{11}}{\sigma_x^2 - \sigma_y^2} \qquad \ldots(5)$$

Eq. (5) gives two values of α. If one is θ, the other is $\frac{\pi}{2} + \theta$.

The corresponding values of p are given by (4). With these values of α and p, (1) gives the equation of the required line. Evidently there are two such lines which are perpendicular.

13.5. Correlation Ratio

Def. Consider the case when corresponding to any given value of x(say x_i) there are more than one value of y(say y_{ij}). Let the pair (x_i, y_{ij}) occur with frequency f_{ij}.

Let
$$\bar{y}_i = \left(\sum_j f_{ij} y_{ij}\right) \Big/ \sum_j f_{ij}$$

Then correlation ratio of y on $x(\eta_{yx})$ is defined by
$$\sum_i \sum_j f_{ij}\,(y_{ij} - \bar{y}_i)^2 = N\,\sigma_y^2 (1 - \eta_{yx}^2)$$
where $\sum_i \sum_j f_{ij} = N$.

Theorem. Show that
$$r^2 \le \eta_{yx}^2 \le 1$$

Proof. Evidently $\eta_{yx}^2 \le 1$.

To prove $r^2 \le \eta_{yx}^2$ first the equation of line of regression of y on x will obtained. Let it be $y = a + bx$.

The unknowns a and b are given by
$$\sum_i \sum_j f_{ij}\, y_{ij} = Na + b \sum_i \sum_j f_{ij} x_i$$
and
$$\sum_i \sum_j f_{ij}\, x_i y_{ij} = a \sum_i \sum_j f_{ij}\, x_i = b \sum_i \sum_j f_{ij}\, x_i^2$$

i.e.,
$$\bar{y} = a + b\bar{x}$$
and
$$\mu_u + \bar{x} \cdot \bar{y} = a\bar{x} + b\left(\sigma_x^2 + \bar{x}^2\right)$$
$$\therefore \qquad b = \frac{\mu_{11}}{\sigma_x^2} \text{ and } a = \bar{y} - \frac{\mu_{11}}{\sigma_x^2}\,\bar{x}$$

$\therefore$ Eq. of line of regression is
$$y - y = \frac{\mu_{11}}{\sigma_x^2}\,(x - \bar{x})$$

Let
$$Y_i = \bar{y} + \frac{\mu_{11}}{\sigma_x^2}\,(x_i - \bar{x})$$

$$\therefore \quad \sum_i \sum_j f_{ij} (y_{ij} - Y_i)^2 = \sum_i \sum_j f_{ij} \left\{ (y_{ij} - \bar{y}) - \frac{\mu_{11}}{\sigma_x^2} (x_i - \bar{x}) \right\}^2$$

$$= N \left(\sigma_y^2 + \frac{\mu_{11}^2}{\sigma_x^2} - 2 \frac{\mu_{11}^2}{\sigma_x^2} \right) = N \sigma_y^2 (1 - r^2)$$

Now $\sum_i \sum_j f_{ij} (y_{ij} - Y_i)^2 \geq \sum_i \sum_j f_{ij} (y_{ij} - \bar{y}_i)^2$

i.e., the sum of square of deviations in any array is least when they are measured from the mean of the array

$$\therefore \quad N \sigma_y^2 (1 - r^2) \geq N \sigma_y^2 (1 - \eta_{yx}^2)$$

which implies $\quad r^2 \leq \eta_{yx}^2$

$$\therefore \quad r^2 \leq \eta_{yx}^2 \leq 1.$$

Note. Similarly as above correlation ratio of x on y (η_{xy}) can be defined and it can be shown that

$$r^2 \leq \eta_{xy}^2 \leq 1.$$

Ex. 13-44. *Show that the correlation ratio of y on x is the ratio of the standard deviation of the weighed means of the arrays of y's (weighed by the corresponding array frequencies) to the standard deviation of all y's of the dist.*

Sol. Let y_{ij} ($j = 1, 2, ...$) be the values of y corresponding to $x = x_i$ and f_{ij} be the frequency of the pair (x_i, y_{ij}).

Now
$$N\sigma_y^2 = \sum_i \sum_j f_{ij} (y_{ij} - \bar{y})^2 \quad \text{where } \bar{y} = \text{A.M. of } y$$

$$= \sum_i \sum_j f_{ij} \{(y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y})\}^2$$

where
$$\bar{y}_i = \left(\sum_j f_{ij} y_{ij} \right) \Big/ \sum_j f_{ij}$$

$$= \sum_i \sum_j f_{ij} (y_{ij} - \bar{y}_i)^2 + \sum_i \sum_j f_{ij} (\bar{y}_i - \bar{y})^2 + 2 \sum_i \sum_j f_{ij} (y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y})$$

$$= N \sigma_y^2 (1 - \eta_{yx}^2) + \sum_i n_i (\bar{y}_i - \bar{y})^2$$

where
$$n_i = \sum_j f_{ij} \qquad \left(\because \sum_j f_{ij} (y_{ij} - \bar{y}_i) = n_i \bar{y}_i - n_i \bar{y}_i = 0 \right)$$

$$\therefore \quad \eta_{yx}^2 = \left\{ \frac{1}{N} \sum_i n_i (\bar{y}_i - \bar{y})^2 \right\} \Big/ \sigma_y^2 = \frac{\sigma_{my}^2}{\sigma_y^2}$$

where $\quad \sigma_{my} = \sqrt{\frac{1}{N} \sum_i n_i (\bar{y}_i - \bar{y})^2}$ is the s.d. of the weighed means of arrays of y's (weighed by the corresponding array frequencies).

Note. For correlation ratio of x on y, $\eta_{xy}^2 = \dfrac{\sigma_{mx}^2}{\sigma_x^2}$.

1. Calculate correlation co-effi

x :	5	15	10
y :	21	14	28

2.

x :	18·8	19·1	17·6
y :	7·8	7·6	7·7

3.

Husband's age (x)	20
Wife's age (y)	14

4.

Husband's age (x)	24	27
Wife's age (y)	18	20

5.

x :	20	18	16	15
y :	12	16	10	14

6.

x :	28	41	40	38
y :	23	34	33	34

7. From the index numbers giv

Months : (*in* 1994)	*May*	*June*	*Ju*
Index no. of prices in Kolkata (x)	169	182	1
Index no. of prices in Mumbai (y)	204	222	2

8. Obtain the co-efficient of c city during the period 1990

Year	1990	1991
Male death rate	33	23
Female death rate	45	31

9. Calculate the correlation co below :

Marks (in Exam. A) x	15	13
Marks (in Exam. B) y	18	16

10. The table below shows the vehicle accidents in a city.

Year	1995	1996
No. of Vehicles with licences ('000)	2·6	2·8
No. of Motor vehicles accidents ('00)	5·9	6·0

$\frac{{}_{11}}{\sigma_x^2}(x_i - \bar{x})\Big\}^2$

$= N\sigma_y^2(1 - r^2)$

when they are measured from the

(η_{xy}) can be defined and it can be

the ratio of the standard deviation
corresponding array frequencies)

esponding to $x = x_i$ and f_{ij} be the

here $\bar{y}$ = A.M. of y

$\bar{y}_i - \bar{y})\}^2$

$\sum_i \sum_j f_{ij}(\bar{y}_i - \bar{y})^2$

$+ 2\sum_i \sum_j f_{ij}(y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y})$

$(\bar{y}_i - \bar{y})^2$

$y_{ij} - \bar{y}_i) = n_i\bar{y}_i - n_i\bar{y}_i = 0\Big)$

$\sigma_y^2 = \frac{\sigma_{my}^2}{\sigma_y^2}$

the s.d. of the weighed means of

uencies).

EXERCISES

1. Calculate correlation co-efficient for the following datas :

x :	5	15	10	20	25	40
y :	21	14	28	7	35	42

[Ans. 0·49]

2.

x :	18·8	19·1	17·6	16·8	18·2	19·5	20·00	21·8	21·9
y :	7·8	7·6	7·7	7·5	7·8	7·2	8·0	7·9	7·8

[Ans. 0·37]

3.

Husband's age (x)	20	30	40	50	60	70	80
Wife's age (y)	14	5	30	32	40	45	65

[Ans. 0·94]

4.

Husband's age (x)	24	27	28	28	29	30	32	33	35	35	40
Wife's age (y)	18	20	22	25	22	28	28	30	27	30	22

[Ans. 0·5]

5.

x :	20	18	16	15	14	12	12	10	8	5
y :	12	16	10	14	12	10	9	8	7	2

[Ans. 0·87]

6.

x :	28	41	40	38	35	33	40	32	36	33
y :	23	34	33	34	30	26	28	31	36	38

[Ans. 0·44]

7. From the index numbers given below, find Karl Pearson's co-efficient of correlation :

Months : (in 1994)	*May*	*June*	*July*	*Aug.*	*Sep.*	*Oct.*	*Now.*	*Dec.*	*Jan.*	*Feb.*
Index no. of prices in Kolkata (x)	169	182	182	192	198	211	227	238	350	253
Index no. of prices in Mumbai (y)	204	222	255	228	231	233	249	266	255	255

[Ans. 0·74]

8. Obtain the co-efficient of correlation between male and female death rates in Delhi city during the period 1990–97.

Year	1990	1991	1992	1993	1994	1995	1996	1997
Male death rate	33	23	24	28	27	28	22	24
Female death rate	45	31	33	40	35	39	32	34

[Ans. 0·97]

9. Calculate the correlation co-efficient between the marks in two examinations given below :

Marks (in Exam. A) x	15	13	17	14	18	12	20	16	18	17	19	21
Marks (in Exam. B) y	18	16	18	15	19	16	18	15	21	17	18	20

[Ans. 0·703]

10. The table below shows the number of vehicles with licences and the number of motor vehicle accidents in a city. Calculate the co-efficient of correlation :

Year	1995	1996	1997	1998	1999	2000	2001	2002
No. of Vehicles with licences ('000)	2·6	2·8	2·9	3·1	3·2	2·3	2·5	1·8
No. of Motor vehicles accidents ('00)	5·9	6·0	6·2	6·2	7·6	7·0	7·4	5·5

[Ans. 0·366]

11. Calculate the co-efficient of correlation between cotton and woollen cloth manufacturer from the following data :

Months	*July*	*Aug.*	*Sep.*	*Oct.*	*Nov.*	*Dec.*
Index no. of cotton cloth manufacturers(x)	103	105	108	106	104	102
Index nos. of woollen cloth manufacturers(y)	75	73	78	71	80	76
	Jan	*Feb.*	*March*	*April*	*May*	*June*
	108	115	118	114	116	120
	68	65	62	60	58	54

[**Ans.** 0·909]

12. Calculate the co-efficient of correlation between the production of rice and its price from the following table :

Production	250	270	278	325	260	510	428	320	440	310
Price	84	50	62	75	90	170	136	65	72	58

[**Ans.** 0·74]

13. Find the correlation co-efficient of datas given below :

$y \to$ / x	30–35	35–40	40–45	45–50	50–55	55–60
80–90	2	3	2	—	—	—
90–100	—	2	5	4	2	—
100–110	—	4	8	5	1	—
110–120	—	—	2	3	1	1
120–130	1	—	—	2	1	1

[**Ans.** 0·43]

14.

$y \to$ / x	18	19	20	21	22	Total
0 – 5	—	—	—	3	1	4
5 – 10	—	—	—	3	2	5
10 – 15	—	—	7	10	—	17
15 – 20	—	5	4	—	—	9
20 – 25	3	2	—	—	—	5
Total	3	7	11	16	3	40

[**Ans.** 0·837]

15. *Ages of daughters (in years)*

Age of mothers (in years)	5—10	10—15	15—20	20—25	25—30	Total
15—25	6	3	—	—	—	9
25—35	3	16	10	—	—	29
35—45	—	10	15	7	—	32
45—55	—	—	7	10	4	21
55—65	—	—	—	4	5	9
Total	9	29	32	21	9	100

[**Ans.** 0·802]

16.

Marks in Sanskrit	0—20
0—20	32
20—40	45
40—60	16
60—80	—
80—100	—
Total	93

17. The following table gives the
Calculate the correlation co

Ages of husbands	10—20
10—20	20
20—30	8
30—40	—
40—50	—

18. Construct examples of at lea
equal to – 1, 0 and + 1.

19. Two independent variates
respectively. Show that the

20. The variables x and y are cc
$ax + by + c = 0$
Show that the correlation cc
are alike and + 1, if signs ar

21. The independent random va
$f(x) = 4ax \quad 0 \le x \le$
$= 0 \quad$ otherwise

Find the correlation co-effic

22. (*a*) Show that
var $(x \pm y)$ = var(x) + var
provided x and y are uncorr
(*b*) Show that
$r_{xy} >$ or < 0 accord

23. If $\bar{x}$ be the A.M. of n indep
$$\text{var}(\bar{x}) = \frac{\sigma^2}{n}.$$

24. If $u = ax + by$, $v = ax - b$
measurements by the same
y is r. If u and v are uncorre

$\sigma_u \alpha_v = 2ab\, \sigma_x \sigma_y$

25. If x_1, x_2, x_3 are three variabl
are uncorrelated, obtain the

26. $x_1, x_2 \ldots. x_n$ are random va
efficient between any two c

...n and woollen cloth manufacturer

...ep.	Oct.	Nov.	Dec.
...08	106	104	102
...78	71	80	76

...rch	April	May	June
...18	114	116	120
...62	60	58	54

[Ans. 0·909]

...e production of rice and its price

...510	428	320	440	310
...170	136	65	72	58

[Ans. 0·74]

...ow :

45–50	50–55	55–60
—	—	—
4	2	—
5	1	—
3	1	1
2	1	1

[Ans. 0·43]

21	22	Total
3	1	4
3	2	5
10	—	17
—	—	9
—	—	5
16	3	40

[Ans. 0·837]

—25	25—30	Total
—	—	9
—	—	29
7	—	32
10	4	21
4	5	9
21	9	100

[Ans. 0·802]

16.

Marks in Sanskrit \ *Marks in History*	0—20	20—40	40—60	60—80	Total
0—20	32	88	15	—	135
20—40	45	436	200	4	685
40—60	16	500	398	25	939
60—80	—	105	532	40	677
80—100	—	8	40	16	64
Total	93	1137	1185	85	2500

[**Ans.** 0·048]

17. The following table gives the ages of husbands and wives at the time of their marriages. Calculate the correlation co-efficient between the ages of husbands and wives :

Ages of husbands \ *Ages of wives*	10—20	20—30	30—40	40—50
10—20	20	26	—	—
20—30	8	14	37	—
30—40	—	4	18	6
40—50	—	—	4	3

[**Ans.** 0·69]

18. Construct examples of at least 5 pairs of observations with co-efficients of correlation equal to -1, 0 and $+1$.

19. Two independent variates x and y have means 5 and 10 and variances 4 and 9 respectively. Show that the variates $u = 3x + 4y$, $v = 3x - y$ are uncorrelated.

20. The variables x and y are connected by the equation

$$ax + by + c = 0$$

Show that the correlation co-efficient between them is -1, if the signs of 'a' and 'b' are alike and $+1$, if signs are different.

21. The independent random variables are defined by

$$f(x) = 4ax \quad 0 \le x \le r \qquad f(y) = 4by \quad 0 \le y \le s$$
$$\;\; = 0 \quad \text{otherwise} \qquad\quad = 0 \quad \text{otherwise}$$

Find the correlation co-efficient between $x + y$ and $x - y$. $\left[\textbf{Ans. } \dfrac{b-a}{b+a}\right]$

22. (*a*) Show that

$$\text{var}(x \pm y) = \text{var}(x) + \text{var}(y)$$

provided x and y are uncorrelated.

(*b*) Show that

$$r_{xy} > \text{or} < 0 \text{ according as } \sigma_{x+y} > \text{ or } < \alpha_{x-y}.$$

23. If $\bar{x}$ be the A.M. of n independent variates $x_1, x_2 \ldots. x_n$ each of s.d. σ, show that

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n}.$$

24. If $u = ax + by$, $v = ax - by$, where x, y represent deviations from the means of two measurements by the same individuals. The co-efficient of correlation between x and y is r. If u and v are uncorrelated, show that

$$\sigma_u \alpha_v = 2ab\, \sigma_x \sigma_y \sqrt{1 - r^2}.$$

25. If x_1, x_2, x_3 are three variables with s.ds. $\sigma_1, \sigma_2, \sigma_3$ respectively. If any two of variables are uncorrelated, obtain the co-efficient of correlation between $x_1 + x_2$, and $x_2 + x_3$.

26. $x_1, x_2 \ldots. x_n$ are random variates each with mean μ and s.d. σ. The correlation co-efficient between any two of them is ρ. Show that

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n} + \left(1-\frac{1}{n}\right)\rho\sigma^2$$

where $\bar{x} = \dfrac{x_1+x_2+\ldots\ldots+x_n}{n}$. Deduce that

$$\rho > -\frac{1}{n-1}.$$

27. x and y are random variates with zero means and unit variances. If

$$r(ax+by,\ bx+ay) = \frac{1+2ab}{a^2+b^2}$$

find $r(x, y)$.

28. Find the co-efficient of correlation between x and y for the following table :

$y \rightarrow$ $x \downarrow$	y_1	y_2	Total
x_1	p_{11}	p_{12}	P
x_2	p_{21}	p_{22}	Q
Total	P'	Q'	1

29. x_1 and x_2 are two variates with variances σ_1^2 and σ_2^2 respectively and r is the co-efficient of correlation between them. Determine the value of the constant k such that $x_1 + kx_2$ and $x_1 + \dfrac{\sigma_1}{\sigma_2}x_2$ are uncorrelated. $\left(\textbf{Ans. } k = -\dfrac{\sigma_1}{\sigma_2}\right)$

30. x_1 and x_2 are independent variables with means 5 and 10 and s.ds. 2 and 3 respectively. Obtain $r(u, v)$ where $u = 3x_1 + 4x_2$, $v = 3x_1 - x_2$. [**Ans.** $r(u, v) = 0$]

31. Let $u = ax + by$, $v = bx - ay$ where x and y represent deviations from their respective means. If the correlation co-efficient between x and y is p and u, v are uncorrelated, show that

(*i*) $ab\,(\sigma_x^2 - \sigma_y^2) = p\sigma_x\sigma_y\,(a^2 - b^2)$

(*ii*) $\sigma_u^2 + \sigma_v^2 = (a^2 + b^2)\,(\sigma_x^2 + \sigma_y^2)$

(See Ex. 13·14)

32. A coin is tossed n times. If x and y denote the number of heads and the number of tails turned up respectively, find $\rho(x, y)$. [ρ is correlation co-efficient].

33. If x, y, z : are three variates each having mean 0, variance 1 and the correlation co-efficient between any two variates is r, show that $r \geq -\dfrac{1}{2}$. What is the corresponding result for n variates? (**Hint.** See Ex. 13·19 and Ex. 26)

34. Two judges in a beauty contest rank the ten competitors in the following orders :

6	4	3	1	2	7	9	8	10	5
4	1	6	7	5	8	10	9	3	2

Calculate the co-efficient of rank correlation. [**Ans.** 0·22]

35. The ranks of the same 15 students in Mathematics and English were as follows, the two numbers within brackets denoting the rank of the same student :
(1, 10), (2, 7), (3, 2), (4, 6), (5, 4), (6, 8), (7, 3), (8, 1), (9, 11), (10, 15), (11, 9), (12, 5), (13, 14), (14, 12), (15, 13)
Find the rank correlation co-efficient. [**Ans.** 0·5]

36. The co-efficient of rank correlation is 0·8. If the sum of the squares of the differences in ranks is 33, find the number of individuals. [**Ans.** 10]

37. The table below shows the
65 63 67 64
68 66 68 65
find co-efficient of rank co

38. For the following data, find
x : 5 15 10
y : 21 14 28

39. Obtain the lines of regressi
x : 1 2 3
y : 9 8 10
Deduce the value of correl
should correspond on the a

40. Determine Karl Pearson's
given in the following table
Exports : 45 46
Imports : 94 96
Obtain also regression equa
[**Ans.** 0·9

41. Mean soil temperature and
above ground) for winter w
Mean soil temp. : 57 42
No. of days : 10 26
Obtain the regression equat

42. Calculate the co-efficient o
two tests :
Student : A B
Test I : 50 54
Test II : 22 25
Also obtain equations of li

43. The following table gives th
and B in an examination :
Marks in A 30—39
30—39 3
40—49 2
50—59 1
60—69 —
Total 6
Calculate the co-efficient o

44. The following regression e
$x = 0{\cdot}8456y + 5{\cdot}4$
$y = 0{\cdot}7326x + 35{\cdot}$
Find the values of (*i*) the co

45. For two variables x and y
$y = ax + b$ and $x = \alpha y + \beta$.

t variances. If

for the following table :

Total
P
Q
1

σ_2^2 respectively and r is the co-
value of the constant k such that

$$\left(\textbf{Ans. } k = -\frac{\sigma_1}{\sigma_2}\right)$$

10 and s.ds. 2 and 3 respectively.
[**Ans.** $r(u, v) = 0$]

deviations from their respective
y is p and u, v are uncorrelated,

r of heads and the number of tails
co-efficient].

ariance 1 and the correlation co-
$-\frac{1}{2}\cdot$ What is the corresponding
Hint. See Ex. 13-19 and Ex. 26)

itors in the following orders :

9	8	10	5
10	9	3	2

[**Ans.** 0·22]

and English were as follows, the
he same student :
, (9, 11), (10, 15), (11, 9), (12, 5),

[**Ans.** 0·5]

of the squares of the differences
[**Ans.** 10]

37. The table below shows the respective heights of 12 fathers and their eldest sons

65	63	67	64	68	62	70	66	68	67	69	71
68	66	68	65	69	66	68	65	71	67	68	70

find co-efficient of rank correlation.

38. For the following data, find the lines of regression :

x :	5	15	10	20	25	30
y :	21	14	28	7	35	43

[**Ans.** $y = 12{\cdot}6 + 0{\cdot}68x$, $x = 9 + 0{\cdot}347y$]

39. Obtain the lines of regression for the following data :

x :	1	2	3	4	5	6	7	8	9
y :	9	8	10	12	11	13	14	16	15

Deduce the value of correlation co-efficient and also obtain an estimate of y which should correspond on the average to $x = 6{\cdot}2$.

[**Ans.** $y = 0{\cdot}95x + 7{\cdot}25$; $x = 0{\cdot}95y - 6{\cdot}4$; 0·95; 13·14]

40. Determine Karl Pearson's co-efficient of correlation between exports and imports given in the following table :

Exports :	45	46	48	50	52	53	51	49	47
Imports :	94	96	98	100	104	105	102	99	97

Obtain also regression equations and standard errors of estimate of x and y.

[**Ans.** 0·99; $y = 1{\cdot}333x + 34{\cdot}111$; $x = 0{\cdot}739y - 24{\cdot}511$, 0·31, 0·42]

41. Mean soil temperature and germination interval (time between sowing and appearance above ground) for winter wheat 1991-96 for 12 places are recorded below :

Mean soil temp. :	57	42	38	42	45	42	44	40	46	44	43	40
No. of days :	10	26	41	29	27	27	19	18	19	31	29	33

Obtain the regression equation of germination interval on mean soil temperature.

[**Ans.** $y = 80{\cdot}752 - 1{\cdot}262x$]

42. Calculate the co-efficient of correlation between the marks secured by 12 students in two tests :

Student :	A	B	C	D	E	F	G	H	I	J	K	L
Test I :	50	54	56	59	60	62	61	65	67	71	71	74
Test II :	22	25	34	28	26	30	33	30	28	34	36	40

Also obtain equations of lines of regression.

[**Ans.** 0·774; $y = 0{\cdot}538x - 3{\cdot}125$, $x = 1{\cdot}115y + 28{\cdot}493$]

43. The following table gives the number of candidates obtaining marks in two subjects A and B in an examination :

	Marks in B				
Marks in A	30—39	40—49	50—59	60—69	Total
30—39	3	1	1	—	5
40—49	2	6	1	2	11
50—59	1	2	2	1	6
60—69	—	1	1	1	3
Total	6	10	5	4	25

Calculate the co-efficient of correlation. Obtain also the lines of regression.

[**Ans.** 0·39; $y = 0{\cdot}43x + 26{\cdot}961$; $x = 0{\cdot}361y + 30{\cdot}225$]

44. The following regression equations are obtained from a correlation table :

$$x = 0{\cdot}8456y + 5{\cdot}45$$
$$y = 0{\cdot}7326x + 35{\cdot}86$$

Find the values of (*i*) the correlation co-efficient (*ii*) the mean of x and y.

45. For two variables x and y with the same mean, the two regression equations are $y = ax + b$ and $x = \alpha y + \beta$. Show that $\dfrac{b}{\beta} = \dfrac{1-a}{1-\alpha}\cdot$ Find also the common mean.

46. Given the regression lines $2y - x = 50$, $3y - 2x = 10$. Show that the estimate of y for $x = 150$ is 100 and the estimate of x for $y = 100$ is 145. Explain the difference.

47. Criticize the following :

$$b_{yx} = 3{\cdot}2 \quad \text{and} \quad b_{xy} = 0{\cdot}8$$

48. Show that $\sin\theta \le (1 - r^2)$ where θ is angle between lines of regression.

Hint. We have $\tan\theta = \dfrac{1-r^2}{r}\,\dfrac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$

Now $\dfrac{\sigma_x^2+\sigma_y^2}{2\sigma_x\sigma_y} - 1 = \dfrac{(\sigma_x-\sigma_y)^2}{2\sigma_x\cdot\sigma_y} \ge 0$

$$\therefore \quad \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2} \le \frac{1}{2}$$

$$\therefore \quad |\tan\theta| \le \frac{1-r^2}{2|r|}$$

$$\therefore \quad |\cot\theta| \ge \frac{2|r|}{1-r^2}$$

$$\therefore \quad 1+\cot^2\theta \ge 1 + \frac{4r^2}{(1-r^2)^2} = \frac{(1+r^2)^2}{(1-r^2)^2}$$

$$\therefore \quad \operatorname{cosec}^2\theta \ge \frac{(1+r^2)^2}{(1-r^2)^2}$$

$$\therefore \quad \sin\theta \le \frac{1-r^2}{1+r^2} \le (1-r^2) \qquad (\because\ 1+r^2 \ge 1)$$

49. If the lines of regression of y on x and x on y are respectively

$$a_1x + b_1y + c_1 = 0$$

and $\quad a_2x + b_2y + c_2 = 0$

show that $\quad a_1b_2 \le a_2b_1$.

50. The two lines of regression are given by $x + 2y = 5$ and $2x + 3y = 8$, calculate

(*i*) the values of $\bar{x}$ and $\bar{y}$

(*ii*) the co-efficient of correlation.

51. Show that, if the random variables x and y have the joint *p.d.f.*

$$f(x, y) = \begin{cases} x+y & 0<x<1,\quad 0<y<1 \\ 0 & \text{elsewhere} \end{cases}$$

then the correlation of x and y is $-\dfrac{1}{11}$.

52. The random variables x and y are jointly normally distributed and u, v are defined by

$$u = x\cos\alpha + y\sin\alpha$$
$$v = y\cos\alpha - x\sin\alpha$$

Show that u and v will be uncorrelated, if

$$\tan 2\alpha = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$$

where ρ = correlation co-efficient between x and y.

□□

Multiple and

14.1. Introduction

Here the theory of correlation inv
aim of the theory of **multiple correlat**
upon a variable not included in the gr
In case, the study of relationship b
remaining variables on these two var
eliminate the entire influence, only the l
the two variables is called **partial cor**
Here only three variables will be

14.2. Notations

Let x_1, x_2, x_3 be the variables. It is
respective means, so that

$$\bar{x}_1 = \bar{x}_2$$

The multiple correlation coeffic
(independent variables) is denoted b
independent variables and figure befor
the multiple correlation between x_2 an
The partial correlation coefficient
after dot refer to variable whose effect
denote the partial correlation between .

14.3. Plane of Regression

The equation of plane of regressio

$$x_1 = a + \ldots$$

where a, $b_{12\cdot3}$ and $b_{13\cdot2}$ are constants. Th
coefficients of x_1 and x_2 for fixed x_3
subscript attached to the b's is the subsc
and the second subscript is that of x t
primary subscripts. The subscript sep
of x which has been left. These are call
Now (1) $\Rightarrow$

$$\bar{x}_1 = a + \ldots$$

$\therefore$ (1) takes the form $x_1 = b_{12\cdot3}$
Here the coefficients b's are to be
$$S = \Sigma(x_1$$
which is the sum of the squares of the r
the variables.

10. Show that the estimate of y for
is 145. Explain the difference.

een lines of regression.

$+ r^2 \geq 1)$

e respectively

$= 5$ and $2x + 3y = 8$, calculate

the joint *p.d.f.*

lly distributed and u, v are defined by

nd y.

14

Multiple and Partial Correlations

14.1. Introduction

Here the theory of correlation involving more than two variables will be discussed. The aim of the theory of **multiple correlation** is to study the joint effect of a group of variables upon a variable not included in the group.

In case, the study of relationship between only two variables is to be made, the effect of remaining variables on these two variables should be eliminated. As it is not possible to eliminate the entire influence, only the linear effect is eliminated. Then the correlation between the two variables is called **partial correlation.**

Here only three variables will be taken.

14.2. Notations

Let x_1, x_2, x_3 be the variables. It is assumed that these denote the deviations from their respective means, so that

$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$$

The multiple correlation coefficient between x_1 (dependent variable) and x_2, x_3 (independent variables) is denoted by $R_{1\cdot23}$. Where figures after dot (.) correspond to independent variables and figure before dot refer to dependent variable. Thus $R_{2\cdot13}$ denote the multiple correlation between x_2 and x_1, x_3 and so on.

The partial correlation coefficient between x_1 and x_2 is denoted by $r_{12\cdot3}$. Where figures after dot refer to variable whose effect has been eliminated or is kept constant. Thus $r_{23\cdot1}$ denote the partial correlation between x_2 and x_3.

14.3. Plane of Regression

The equation of plane of regression of x_1 on x_2, x_3 is of the form

$$x_1 = a + b_{12\cdot3}\, x_2 + b_{13\cdot2}\, x_3 \qquad \text{...(1)}$$

where a, $b_{12\cdot3}$ and $b_{13\cdot2}$ are constants. The quantities $b_{12\cdot3}$ and $b_{13\cdot2}$ are called partial regression coefficients of x_1 and x_2 for fixed x_3 and of x_1 on x_3 for fixed x_2 respectively. The first subscript attached to the b's is the subscript of the letter on the left (the dependent variable) and the second subscript is that of x to which it is attached. These subscripts are called **primary subscripts**. The subscript separated from the primary subscripts by a dot (.) is that of x which has been left. These are called **secondary subscripts.**

Now (1) $\Rightarrow$

$$\bar{x}_1 = a + b_{12\cdot3}\, \bar{x}_2 + b_{13\cdot2}\, \bar{x}_3 \Rightarrow a = 0$$

$\therefore$ (1) takes the form $x_1 = b_{12\cdot3}\, x_2 + b_{13\cdot2}\, x_3$ ...(2)

Here the coefficients b's are to be obtained so as to minimize

$$S = \Sigma(x_1 - b_{12\cdot3}\, x_2 - b_{13\cdot2}\, x_3)^2$$

which is the sum of the squares of the residuals, the summation is over the given values of the variables.

The normal equations are

$$0 = \frac{\partial S}{\partial b_{12\cdot3}} = -2\Sigma x_2 (x_1 - b_{12\cdot3}\, x_2 - b_{13\cdot2}\, x_3)$$

$$\Rightarrow \quad \Sigma x_1 x_2 = b_{12\cdot3}\, \Sigma x_2^2 + b_{13\cdot2}\, \Sigma x_2 x_3$$

$$\Rightarrow \quad nr_{12}\, \sigma_1 \sigma_2 = b_{12\cdot3}\, n\sigma_2^2 + b_{13\cdot2}\, nr_{23}\, \sigma_2\sigma_3$$

$$\Rightarrow \quad r_{12}\sigma_1 = b_{12\cdot3}\, \sigma_2 + b_{13\cdot2}\, r_{23}\, \sigma_3 \qquad \text{...(3)}$$

and

$$0 = \frac{\partial S}{\partial b_{13\cdot2}} = -2\Sigma x_3 (x_1 - b_{12\cdot3}\, x_2 - b_{13\cdot2}\, x_3)$$

$$\Rightarrow \quad \Sigma x_1 x_3 = b_{12\cdot3}\, \Sigma x_2 x_3 + b_{13\cdot2}\, \Sigma x_3^2$$

$$\Rightarrow \quad r_{13}\, \sigma_1 = b_{12\cdot3}\, \sigma_2\, r_{23} + b_{13\cdot2}\, \sigma_3 \qquad \text{...(4)}$$

Solving (3) and (4)

$$b_{12\cdot3} = \frac{\begin{vmatrix} r_{12}\,\sigma_1 & r_{23}\sigma_3 \\ r_{13}\,\sigma_1 & \sigma_3 \end{vmatrix}}{\begin{vmatrix} \sigma_2 & r_{23}\,\sigma_3 \\ r_{23}\sigma_2 & \sigma_3 \end{vmatrix}} = \frac{\sigma_1}{\sigma_2} \frac{\begin{vmatrix} r_{12} & r_{23} \\ r_{13} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}}$$

and

$$b_{13\cdot2} = \frac{\sigma_1}{\sigma_3} \frac{\begin{vmatrix} 1 & r_{12} \\ r_{23} & r_{13} \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}}$$

For convenience and simplicity, let

$$\omega = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix}$$

and observe

$$r_{11} = r_{22} = r_{33} = 1$$

$$r_{12} = r_{21},\ r_{13} = r_{31},\ r_{23} = r_{32}$$

Let ω_{ij} = cofactor of (i, j)th place.

Then

$$\omega_{11} = \begin{vmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{vmatrix} = \begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}$$

$$\omega_{12} = -\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & r_{33} \end{vmatrix} = -\begin{vmatrix} r_{12} & r_{23} \\ r_{13} & 1 \end{vmatrix}$$

$$\omega_{13} = \begin{vmatrix} r_{21} & r_{22} \\ r_{31} & r_{32} \end{vmatrix} = -\begin{vmatrix} 1 & r_{12} \\ r_{23} & r_{13} \end{vmatrix}$$

$$\therefore \quad b_{12\cdot3} = -\frac{\sigma_1}{\sigma_2} \cdot \frac{\omega_{12}}{\omega_{11}},\quad b_{13\cdot2} = -\frac{\sigma_1}{\sigma_3} \frac{\omega_{13}}{\omega_{11}} \qquad \text{...(5)}$$

Substituting these values in (2), Eq. of plane of regression of x_1 on x_2, x_3 is

$$x_1 = -\frac{\sigma_1}{\sigma_2} \frac{\omega_{12}}{\omega_{11}}\, x_2 - \frac{\sigma_1}{\sigma_3} \frac{\omega_{13}}{\omega_{11}}\, x_3$$

i.e.,

$$\frac{x_1}{\sigma_1}\, \omega_{11} + \frac{x_2}{\sigma_2}\, \omega_{12} + \frac{x_3}{\sigma_3}\, \omega_{13} = 0. \qquad \text{...(6)}$$

Similarly eqs. of planes of regression of x_2 on x_1, x_3 and x_3 on x_1, x_2 respectively are

$$\frac{x_1}{\sigma_1}\, \omega_2$$

and

$$\frac{x_1}{\sigma_1}\, \omega_3$$

Remark. Eliminating $b_{12\cdot3}$ and b_1

$$\begin{vmatrix} x_1 \\ r_{12}\sigma_1 \\ r_{13}\sigma_1 \end{vmatrix}$$

i.e.,

$$\begin{vmatrix} \frac{x_1}{\sigma_1} \\ r_{12} \\ r_{13} \end{vmatrix}$$

which is the eq. of plane of regression

Remark. (1) x_1, x_2, x_3 are also cons
Then value of x_1 as estimated by plane
denoted by $\varepsilon_{1\cdot23}$. Thus

$$\epsilon_{1\cdot23} = b_{12\cdot3}$$

Similarly

$$\epsilon_{2\cdot13} = b_{21\cdot3}$$

$$\epsilon_{3\cdot12} = b_{31\cdot2}$$

The difference $x_1 - \epsilon_{1\cdot23}$ is the resi
Thus $x_{1\cdot23} = x_1 -$
$= x_1 -$
Similarly $x_{2\cdot13} = x_2 -$
$x_{3\cdot12} = x_3 -$

(2) In a quantity the subscripts be
those after dot are called **secondary sub**
any order but the order of primary subs
left refers to the dependent variable and
The order of the quantity is detern
Thus $x_{1\cdot23}$ is of order two, $b_{12\cdot3}$ is order

Ex. 14-1. *Using the following data*

$\bar{x}_1 = 40$
$\sigma_1 = 3$
$r_{12} = 0{\cdot}4$

find the equation of plane of regression
30, $x_3 = 40$.

Sol. We have

$$\omega = \begin{vmatrix} 1 \\ r_{12} \\ r_{13} \end{vmatrix}$$

Eq. of plane of regression of x_2 on

$$\frac{x_1}{\sigma_1}\, \omega_{21}$$

$$\frac{x_1}{\sigma_1}\omega_{21} + \frac{x_2}{\sigma_2}\omega_{22} + \frac{x_3}{\sigma_3}\omega_{23} = 0. \qquad ...(7)$$

and
$$\frac{x_1}{\sigma_1}\omega_{31} + \frac{x_2}{\sigma_2}\omega_{32} + \frac{x_3}{\sigma_3}\omega_{33} = 0. \qquad ...(8)$$

Remark. Eliminating $b_{12\cdot3}$ and $b_{13\cdot2}$ between (2), (3) and (4).

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ r_{12}\sigma_1 & \sigma_2 & r_{23}\sigma_3 \\ r_{13}\sigma_1 & \sigma_2 r_{23} & \sigma_3 \end{vmatrix} = 0$$

i.e.,
$$\begin{vmatrix} \dfrac{x_1}{\sigma_1} & \dfrac{x_2}{\sigma_2} & \dfrac{x_3}{\sigma_3} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} = 0$$

which is the eq. of plane of regression of x_1 on x_2, x_3 in determinant form.

Remark. (1) x_1, x_2, x_3 are also considered as the observed values of variates respectively. Then value of x_1 as estimated by plane of regression is $b_{12\cdot3}\,x_2 + b_{13\cdot2}\,x_3$. Let this value be denoted by $\varepsilon_{1\cdot23}$. Thus

$$\in_{1\cdot23} = b_{12\cdot3}\,x_2 + b_{13\cdot2}\,x_3$$

Similarly

$$\in_{2\cdot13} = b_{21\cdot3}\,x_1 + b_{23\cdot1}\,x_3$$
$$\in_{3\cdot12} = b_{31\cdot2}\,x_1 + b_{32\cdot1}\,x_2$$

The difference $x_1 - \in_{1\cdot23}$ is the residual of x_1. It is denoted by $x_{1\cdot23}$.

Thus
$$\begin{aligned} x_{1\cdot23} &= x_1 - \in_{1\cdot23} \\ &= x_1 - b_{12\cdot3}\,x_2 - b_{13\cdot2}\,x_3 \end{aligned}$$

Similarly
$$\begin{aligned} x_{2\cdot13} &= x_2 - b_{21\cdot3}\,x_1 - b_{23\cdot1}\,x_3 \\ x_{3\cdot12} &= x_3 - b_{31\cdot2}\,x_1 - b_{32\cdot1}\,x_2. \end{aligned}$$

(2) In a quantity the subscripts before dot (.) are known as **primary subscripts** and those after dot are called **secondary subscripts**. The secondary subscripts can be written in any order but the order of primary subscripts is important. First primary subscript from the left refers to the dependent variable and other to independent variable.

The order of the quantity is determined by the number of secondary subscripts in it. Thus $x_{1\cdot23}$ is of order two, $b_{12\cdot3}$ is order one and so on.

Ex. 14-1. *Using the following data*

$\bar{x}_1 = 40$	$\bar{x}_2 = 70$	$\bar{x}_3 = 90$
$\sigma_1 = 3$	$\sigma_2 = 6$	$\sigma_3 = 7$
$r_{12} = 0{\cdot}4$	$r_{23} = 0{\cdot}5$	$r_{13} = 0{\cdot}6$

find the equation of plane of regression of x_2 on x_1 and x_3. Also find the value of x_2 for $x_1 = 30$, $x_3 = 40$.

Sol. We have

$$\omega = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0{\cdot}4 & 0{\cdot}6 \\ 0{\cdot}4 & 1 & 0{\cdot}5 \\ 0{\cdot}6 & 0{\cdot}5 & 1 \end{vmatrix}$$

Eq. of plane of regression of x_2 on x_1 and x_3 is

$$\frac{x_1}{\sigma_1}\omega_{21} + \frac{x_2}{\sigma_2}\omega_{22} + \frac{x_3}{\sigma_3}\omega_{23} = 0$$

Here x_1, x_2, x_3 are with zero means. So these are to be replaced by

$$x_1-\bar{x}_1, x_2-\bar{x}_2, x_3-\bar{x}_3$$

respectively.

$\therefore$ Eq. of plane of regression of x_2 on x_1 and x_3 is

$$\frac{(x_1-\bar{x}_1)}{\sigma_1}\omega_{21}+\frac{(x_2-\bar{x}_2)}{\sigma_2}\omega_{22}+\frac{(x_3-\bar{x}_3)}{\sigma_3}\omega_{23}=0 \qquad ...(1)$$

Now

$$\omega_{21} = -\begin{vmatrix} 0\cdot4 & 0\cdot6 \\ 0\cdot5 & 1 \end{vmatrix} = 0\cdot3-0\cdot4=-0\cdot1$$

$$\omega_{22} = \begin{vmatrix} 1 & 0\cdot6 \\ 0\cdot6 & 1 \end{vmatrix} = 1-0\cdot36=0\cdot64$$

$$\omega_{23} = -\begin{vmatrix} 1 & 0\cdot4 \\ 0\cdot6 & 0\cdot5 \end{vmatrix} = 0\cdot24-0\cdot5=-0\cdot26$$

$\therefore$ Substituting values in (1), eq. of plane of regression of x_2 on x_1 and x_3 is

$$\frac{(x_1-40)}{3}(-0\cdot1)+\frac{(x_2-70)}{6}(0\cdot64)+\frac{(x_3-90)}{7}(-0\cdot26)=0$$

$$-0\cdot03\,(x_1-40)+0\cdot11(x_2-70)-0\cdot04\,(x_3-90)=0$$

$$-0\cdot03\,x_1+0\cdot11x_2-0\cdot04\,x_3-2\cdot9=0$$

$\therefore$ $0\cdot03\,x_1-0\cdot11x_2+0\cdot04\,x_3+2\cdot9=0.$

Put $x_1=30,\ x_3=40$

$\therefore$ $0\cdot9-0\cdot11\,x_2+1\cdot6+2\cdot9=0$

$0\cdot11x_2=5\cdot4 \qquad \therefore\ x_2=49\cdot09.$

14.4. Properties of Residuals

(*i*) In the derivation of plane of regression of x_1 on x_2, x_3, normal equations are

$$\Sigma x_2\,(x_1-b_{12\cdot3}\,x_2-b_{13\cdot2}\,x_3)=0$$

and

$$\Sigma x_3\,(x_1-b_{12\cdot3}\,x_2-b_{13\cdot2}\,x_3)=0$$

These equations $\Rightarrow$ $\Sigma x_2\,x_{1\cdot23}=0$ and $\Sigma x_3\,x_{1\cdot23}=0$

Similarly, $\Sigma x_1\,x_{2\cdot13} = 0$ and $\Sigma x_3\,x_{2\cdot13}=0$

$\Sigma x_1x_{3\cdot12} = 0 = \Sigma x_2x_{3\cdot12}$

Thus "the sum of the product of any residual of zero order with any other higher order residual (having the subscripts of the former as one of its secondary subscripts) is zero."

(2)
$$\begin{aligned}\Sigma x_{1\cdot2}\,x_{1\cdot23} &= \Sigma(x_1-b_{12}x_2)\,x_{1\cdot23}\\ &= \Sigma x_1x_{1\cdot23}-b_{12}\Sigma x_2\,x_{1\cdot23}\\ &= \Sigma x_1x_{1\cdot23} \qquad \{\text{using property (1)}\}\end{aligned}$$

Similarly $\Sigma x_{1\cdot3}x_{1\cdot23} = \Sigma x_1x_{1\cdot23}$

Also
$$\begin{aligned}\Sigma x_{1\cdot23}{}^2 &= \Sigma x_{1\cdot23}\,x_{1\cdot23}\\ &= \Sigma(x_1-b_{12\cdot3}\,x_2-b_{13\cdot2}\,x_3)\,x_{1\cdot23}\\ &= \Sigma x_1\,x_{1\cdot23}-b_{12\cdot3}\,\Sigma x_2\,x_{1\cdot23}-b_{13\cdot2}\,\Sigma x_3\,x_{1\cdot23}\\ &= \Sigma x_1\,x_{1\cdot23}\end{aligned}$$

Thus $\Sigma x_{1\cdot2}\,x_{1\cdot23} = \Sigma x_{1\cdot3}x_{1\cdot23}=\Sigma x_{1\cdot23}{}^2=\Sigma x_1\,x_{1\cdot23}$

Similarly $\Sigma x_{2\cdot1}\,x_{2\cdot13} = \Sigma x_{2\cdot3}x_{2\cdot13}=\Sigma x_{2\cdot13}{}^2= \Sigma x_2\,x_{2\cdot13}$

$\Sigma x_{3\cdot1}\,x_{3\cdot12} = \Sigma x_{3\cdot2}\,x_{3\cdot12}=\Sigma x_{3\cdot12}{}^2= \Sigma x_3\,x_{3\cdot12}.$

Thus, "in the sum of product of any two residuals in which all the secondary subscripts of first occur among the secondary subscripts of the second, all the secondary subscripts of the first can be omitted".

(3)
$$\begin{aligned}\Sigma x_{1\cdot2}\,x_{3\cdot12} &= \Sigma(x_1-b_{12}\,x_2)\,x_{3\cdot12}\\ &= \Sigma x_1\,x_{3\cdot12}-b_{12}\,\Sigma x_2\,x_{3\cdot12}=0\end{aligned}$$

Similarly $\Sigma x_{1\cdot 3}\, x_{2\cdot 13} = 0,\ \Sigma x_{2\cdot 1}\, x_{3\cdot 12} = 0,\ \Sigma x_{2\cdot 3}\, x_{1\cdot 23} = 0$ etc.

Thus "the sum of the product of two residuals is zero provided all the subscripts (primary as well as secondary) of one occur among the secondary subscripts of the other".

14.5. Multiple Correlation Coefficient

Multiple correlation coefficient of x_1 on x_2, x_3 is the simple correlation coefficient between x_1 and its value is given by the plane of regression of x_1 on x_2, x_3 *viz.*,

$$\in_{1\cdot 23} = b_{12\cdot 3}\, x_2 + b_{13\cdot 2}\, x_3$$

Let N be the total number of observations for each variate.

Now

$$\begin{aligned} \overline{\in}_{1\cdot 23} &= b_{12\cdot 3}\, \bar{x}_2 + b_{13\cdot 2}\, \bar{x}_3 \\ &= 0 \end{aligned}$$

$\therefore$

$$\begin{aligned} \text{Var}\,(\in_{1\cdot 23}) &= \frac{1}{N}\, \Sigma\, \{\in_{1\cdot 23}\}^2 \\ &= \frac{1}{N}\, \Sigma \{b_{12\cdot 3}\, x_2 + b_{13\cdot 2}\, x_3\}^2 \\ &= b_{12\cdot 3}^2 \cdot \frac{\Sigma\left(x_2^2\right)}{N} + b_{13\cdot 2}^2\, \frac{\Sigma\left(x_3^2\right)}{N} + 2 b_{12\cdot 3}\, b_{13\cdot 2} \cdot \frac{\Sigma (x_2\, x_3)}{N} \\ &= b_{12\cdot 3}^2\, \sigma_2^2 + b_{13\cdot 2}^2\, \sigma_3^2 + 2 b_{12\cdot 3}\, b_{13\cdot 2}\ \text{cov}\,(x_2, x_3) \\ &= \left(-\frac{\sigma_1}{\sigma_2}\,\frac{\omega_{12}}{\omega_{11}}\right)^2 \sigma_2^2 + \left(-\frac{\sigma_1}{\sigma_3}\,\frac{\omega_{13}}{\omega_{11}}\right)^2 \sigma_3^2 \\ &\quad + 2\left(-\frac{\sigma_1}{\sigma_2}\,\frac{\omega_{12}}{\omega_{11}}\right)\left(-\frac{\sigma_1}{\sigma_3}\,\frac{\omega_{13}}{\omega_{11}}\right) r_{23}\, \sigma_2\, \sigma_3 \\ &= \frac{\sigma_1^2}{\omega_{11}^2}\left(\omega_{12}^2 + \omega_{13}^2 + 2 r_{23}\, \omega_{12}\omega_{13}\right) \\ &= \frac{\sigma_1^2}{\omega_{11}^2}\, \{(r_{13} r_{23} - r_{12}\}^2 + (r_{12}\, r_{23} - r_{13})^2 \\ &\qquad + 2 r_{23}\, (r_{13} r_{23} - r_{12})\, (r_{12}\, r_{23} - r_{13})\} \\ &= \frac{\sigma_1^2}{\omega_{11}^2} \left\{\left(r_{12}^2 + r_{13}^2 - 2 r_{12}\, r_{13}\, r_{23}\right)\left(1 - r_{23}^2\right)\right\} \\ &= \frac{\sigma_1^2}{\omega_{11}} \left\{r_{12}^2 + r_{13}^2 - 2 r_{12}\, r_{13}\, r_{23}\right\} \\ &= \frac{\sigma_1^2}{\omega_{11}} \{\omega_{11} - \omega\} = \sigma_1^2 \left\{1 - \frac{\omega}{\omega_{11}}\right\} \qquad \ldots(1) \end{aligned}$$

where

$$\begin{aligned} \omega &= \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} \\ &= 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23}. \end{aligned}$$

$$\text{cov}\,(x_1, \in_{1\cdot 23}) = \frac{1}{N}\, \Sigma x_1\, \in_{1\cdot 23}$$

$$= \frac{1}{N}\,\Sigma x_1\,(b_{12\cdot3}\,x_2 + b_{13\cdot2}\,x_3)$$
$$= b_{12\cdot3}\,\text{cov}\,(x_1, x_2) + b_{13\cdot2}\,\text{cov}\,(x_1, x_3)$$
$$= \left(-\frac{\sigma_1}{\sigma_2}\,\frac{\omega_{12}}{\omega_{11}}\right)\sigma_1\sigma_2 r_{12} + \left(-\frac{\sigma_1}{\sigma_3}\,\frac{\omega_{13}}{\omega_{11}}\right)\sigma_1\sigma_3 r_{13}$$
$$= -\frac{\sigma_1^2}{\omega_{11}}\,\{\omega_{12}r_{12} + \omega_{13}r_{13}\}$$
$$= -\frac{\sigma_1^2}{\omega_{11}}\,\{(r_{13}r_{23} - r_{12})r_{12} + (r_{12}r_{23} - r_{13})r_{13}\}$$
$$= \frac{\sigma_1^2}{\omega_{11}}\left\{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}\right\}$$
$$= \sigma_1{}^2\left\{1 - \frac{\omega}{\omega_{11}}\right\} \qquad \text{...(2)}$$

$$\therefore \qquad R_{1\cdot23}^2 = 1 - \frac{\omega}{\omega_{11}}$$
$$= \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}.$$

Remark. (1) and (2) $\Rightarrow$

$$\text{Cov}(x_1, \in_{1\cdot23}) = \text{Var}\,(\in_{1\cdot23}) \geq 0$$
$$\therefore \qquad R_{1\cdot23} \geq 0$$

Also since $R_{1\cdot23}$ is simple correlation coefficient, $R_{1\cdot23} \leq 1$

$$\therefore \qquad 0 \leq R_{1\cdot23} \leq 1.$$

Ex. 14-2. *Three variables have in pairs simple correlation coefficients given by* $r_{12} = -0{\cdot}8 \quad r_{13} = 0{\cdot}7 \quad r_{23} = -0{\cdot}9$.

Find the multiple correlation coefficient $R_{1\cdot23}$ of x_1 on x_2 and x_3.

Sol. We have

$$\omega = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0{\cdot}8 & -0{\cdot}7 \\ 0{\cdot}8 & 1 & -0{\cdot}9 \\ -0{\cdot}7 & -0{\cdot}9 & 1 \end{vmatrix} = {\cdot}07$$

$$\omega_{11} = \begin{vmatrix} 1 & -0{\cdot}9 \\ -0{\cdot}9 & 1 \end{vmatrix} = 1 - 0{\cdot}81 = 0{\cdot}19$$

$$\therefore \qquad R_{1\cdot23}^2 = 1 - \frac{\omega}{\omega_{11}} = 1 - \frac{{\cdot}07}{{\cdot}19} = \frac{12}{19} = 0{\cdot}63$$
$$\therefore \qquad R_{1\cdot23} = 0{\cdot}8.$$

Ex. 14-3. *Show that* $R_{1\cdot23}^2 = 1 - \dfrac{\sigma_{1\cdot23}^2}{\sigma_1^2}$ *where* $\sigma_{1\cdot23}$ denotes the s.d. of $x_{1\cdot23}$.

Sol. We have
$$\in_{1\cdot23} = b_{12\cdot3}\,x_2 + b_{13\cdot2}\,x_3 = x_1 - x_{1\cdot23}$$
$$\therefore \qquad \text{Var}\,(\in_{1\cdot23}) = \frac{1}{N}\,\Sigma\{x_1 - x_{1\cdot23})^2$$
$$= \frac{1}{N}\,\Sigma\left\{\sigma_1^2 + x_{1\cdot23}^2 - 2x_1\,x_{1\cdot23}\right\}$$

$= \frac{1}{N}$

$= \sigma_1^2$

$= \sigma_1^2$

Cov $(x_1, \in_{1\cdot23}) = \frac{1}{N}$

$= \frac{1}{N}$

$= \sigma_1^2$

$\therefore \qquad R_{1\cdot23}^2 = \{$

$= 1 -$

Ex. 14-4. *Show that* $\sigma_{1\cdot32}^2 = \sigma_1^2$

Sol. We have $\sigma_{1\cdot32}^2 = \frac{1}{N}$

$= \frac{1}{N}$

$= \frac{1}{N}$

$= \frac{1}{N}$

$= \sigma_1$

$= \sigma_1$

$= \sigma_1$

$= \frac{\sigma}{\omega}$

Since ω_{ij}'s are cofactor in ω,

$r_{11}\,\omega$

$\therefore \qquad \sigma_{1\cdot32}^2 = \sigma$

14.6. Partial Correlation Coefficie

As already defined, the partial c between x_1 and x_2 after the linear eff

Now linear effect of x_3 on x_1 as

linear effect of x_3 on x_2 is $b_{23} = r_{23}\,\frac{\sigma}{\sigma}$

$$= \frac{1}{N}\Sigma x_1^2 + \frac{1}{N}\Sigma x_{1\cdot 23}^2 - 2\cdot\frac{1}{N}\Sigma x_1\, x_{1\cdot 23}$$

$$= \sigma_1^2 + \sigma_{1\cdot 23}^2 - 2\frac{1}{N}\Sigma x_{1\cdot 23}^2$$

$$= \sigma_1^2 - \sigma_{1\cdot 23}^2$$

$$\text{Cov}\,(x_1, \in_{1\cdot 23}) = \frac{1}{N}\Sigma x_1(x_1 - x_{1\cdot 23})$$

$$= \frac{1}{N}\Sigma x_1^2 - \frac{1}{N}\Sigma x_1\, x_{1\cdot 23}$$

$$= \sigma_1^2 - \frac{1}{N}\Sigma x_{1\cdot 23}^2 = \sigma_1^2 - \sigma_{1\cdot 23}^2$$

$$\therefore \quad R_{1\cdot 23}^2 = \left\{\frac{\text{Cov}(x_1, \in_{1\cdot 23})}{(\text{s.d. of } x_1)(s.d.of \in_{1\cdot 23})}\right\}^2$$

$$= 1 - \frac{\sigma_{1\cdot 23}^2}{\sigma_1^2} = 1 - \frac{\omega}{\omega_{11}}. \qquad \text{(See Ex. 14-4)}$$

Ex. 14-4. *Show that* $\sigma_{1\cdot 32}^2 = \sigma_1^2\,\dfrac{\omega}{\omega_{11}}$.

Sol. We have

$$\sigma_{1\cdot 32}^2 = \frac{1}{N}\Sigma x_{1\cdot 32}^2$$

$$= \frac{1}{N}\Sigma x_{1\cdot 32}\, x_{1\cdot 32}$$

$$= \frac{1}{N}\Sigma x_1\, x_{1\cdot 32}$$

$$= \frac{1}{N}\Sigma x_1\,(x_1 - b_{12\cdot 3}\, x_2 - b_{13\cdot 2}\, x_3)$$

$$= \sigma_1{}^2 - b_{1\cdot 23}\,\text{Cov}\,(x_1, x_2) - b_{13\cdot 2}\,\text{Cov}\,(x_1, x_3)$$

$$= \sigma_1{}^2 - r_{12}\,\sigma_1\,\sigma_2\left(-\frac{\sigma_1}{\sigma_2}\frac{\omega_{12}}{\omega_{11}}\right) - r_{13}\,\sigma_1\sigma_3\left(-\frac{\sigma_1}{\sigma_3}\frac{\omega_{13}}{\omega_{11}}\right)$$

$$= \sigma_1^2\left(1 + r_{12}\frac{\omega_{12}}{\omega_{11}} + r_{13}\frac{\omega_{13}}{\omega_{11}}\right)$$

$$= \frac{\sigma_1^2}{\omega_{11}}\,(r_{11}\,\omega_{11} + r_{12}\,\omega_{12} + r_{13}\,\omega_{13})$$

Since ω_{ij}'s are cofactor in ω,

$$r_{11}\,\omega_{11} + r_{12}\,\omega_{12} + r_{13}\,\omega_{13} = \omega$$

$$\therefore \quad \sigma_{1\cdot 32}^2 = \sigma_1^2\,\frac{\omega}{\omega_{11}}.$$

14.6. Partial Correlation Coefficient

As already defined, the partial correlation between x_1 and x_2 is the simple correlation between x_1 and x_2 after the linear effect of x_3 on them has been eliminated.

Now linear effect of x_3 on x_1 as indicated by regression of x_1 on x_3 is $b_{13} = r_{13}\dfrac{\sigma_1}{\sigma_3}$ and linear effect of x_3 on x_2 is $b_{23} = r_{23}\dfrac{\sigma_2}{\sigma_3}$.

$$\therefore \qquad x_{1\cdot 3} = x_1 - r_{13}\frac{\sigma_1}{\sigma_3}x_3$$

and
$$x_{2\cdot 3} = x_2 - r_{23}\frac{\sigma_2}{\sigma_3}x_3$$

are parts of x_1, x_2 respectively, which remain after the elimination of linear effect of x_3.

Thus partial correlation coefficient between x_1, x_2 is the simple correlation coefficient between $x_{1\cdot 3}$, $x_{2\cdot 3}$.

Now
$$\begin{aligned}
\text{Cov}(x_{1\cdot 3}, x_{2\cdot 3}) &= \frac{1}{N}\Sigma x_{1\cdot 3}\, x_{2\cdot 3} \qquad \{\because\ \bar{x}_{1\cdot 3} = \bar{x}_{2\cdot 3} = 0\}\\
&= \frac{1}{N}\sum\left(x_1 - r_{13}\frac{\sigma_1}{\sigma_3}x_3\right)\left(x_2 - r_{23}\frac{\sigma_2}{\sigma_3}x_3\right)\\
&= \frac{1}{N}\sum\left\{x_1x_2 - r_{13}\frac{\sigma_1}{\sigma_3}x_3x_2 - r_{23}\frac{\sigma_2}{\sigma_3}x_1x_3 + r_{13}r_{23}\frac{\sigma_1}{\sigma_3}\cdot\frac{\sigma_2}{\sigma_3}x_3^2\right\}\\
&= \text{Cov}(x_1, x_2) - r_{13}\frac{\sigma_1}{\sigma_3}\text{Cov}(x_2, x_3) - r_{23}\frac{\sigma_2}{\sigma_3}\text{Cov}(x_1, x_3) + r_{13}r_{23}\frac{\sigma_1\sigma_2}{\sigma_3^2}\cdot\sigma_3^2\\
&= \sigma_1\sigma_2(r_{12} - r_{13}r_{23})
\end{aligned}$$

$$\begin{aligned}
\text{Var}(x_{1\cdot 3}) &= \frac{1}{N}\Sigma x_{1\cdot 3}^2\\
&= \frac{1}{N}\Sigma x_{1\cdot 3}\, x_{1\cdot 3}\\
&= \frac{1}{N}\Sigma x_1\, x_{1\cdot 3}\\
&= \frac{1}{N}\Sigma x_1\left[x_1 - r_{13}\frac{\sigma_1}{\sigma_3}x_3\right] = \sigma_1^2\left(1 - r_{13}^2\right)
\end{aligned}$$

Similarly
$$\text{Var}(x_{2\cdot 3}) = \sigma_2^2(1 - r_{23}^2)$$

$$\begin{aligned}
\therefore \qquad r_{12\cdot 3} &= \frac{\text{Cov}(x_{1\cdot 3}, x_{2\cdot 3})}{(\text{s.d. of } x_{1\cdot 3})(\text{s.d. of } x_{2\cdot 3})}\\
&= \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{-\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}}
\end{aligned}$$

Remark. If $r_{12\cdot 3} = 0$, then $r_{12} = r_{13}\, r_{23}$.

$\therefore$ If x_3 is correlated with x_1, x_2 both *i.e.*, $r_{23} \neq 0$, $r_{13} \neq 0$, then

$$r_{12} \neq 0$$

$\Rightarrow$ x_1, x_2 are not uncorrelated.

$\therefore$ x_1, x_2 are correlated even though they are uncorrelated after the effect of x_3 is eliminated.

This is because x_1, x_2 carry the effect of x_3 on them.

mination of linear effect of x_3.
the simple correlation coefficient

$\{\because \bar{x}_{1\cdot3} = \bar{x}_{2\cdot3} = 0\}$

$\Big)\left(x_2 - r_{23}\dfrac{\sigma_2}{\sigma_3}x_3\right)$

$x_3x_2 - r_{23}\dfrac{\sigma_2}{\sigma_3}x_1x_3$

$\Big\}$

$\text{Cov}(x_2, x_3) - r_{23}\dfrac{\sigma_2}{\sigma_3}\text{Cov}(x_1, x_3)$

$+ r_{13}r_{23}\dfrac{\sigma_1\sigma_2}{\sigma_3^2}\cdot\sigma_3^2$

$_3\Big] = \sigma_1^2\left(1 - r_{13}^2\right)$

$_{2\cdot3})$

$\dfrac{-\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}}$

$_3 \neq 0$, then

correlated after the effect of x_3 is

Ex. 14-5. *Show that*

$$(i)\quad r_{23\cdot1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{(1-r_{21}^2)(1-r_{31}^2)}}$$

$$(ii)\quad r_{13\cdot2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}}.$$

Sol. (*i*) We have

$$\begin{aligned}
0 &= \Sigma x_{2\cdot13}\,x_{3\cdot1} \\
&= \Sigma(x_2 - b_{21\cdot3}\,x_1 - b_{23\cdot1}\,x_3)\,x_{3\cdot1} \\
&= \Sigma x_2\,x_{3\cdot1} - b_{21\cdot3}\,\Sigma x_1\,x_{3\cdot1} - b_{23\cdot1}\,\Sigma x_3\,x_{3\cdot1} \\
&= \Sigma x_{2\cdot1}\,x_{3\cdot1} - b_{23\cdot1}\,\Sigma x_{3\cdot1}{}^2
\end{aligned}$$

$$\Rightarrow \qquad b_{23\cdot1} = \frac{\Sigma x_{2\cdot1}\,x_{3\cdot1}}{\Sigma x_{3\cdot1}^2} = \frac{\text{Cov}(x_{2\cdot1}, x_{3\cdot1})}{\sigma_{3\cdot1}^2}$$

$$= r_{23\cdot1}\frac{\sigma_{2\cdot1}}{\sigma_{3\cdot1}}$$

Similarly $$b_{32\cdot1} = r_{23\cdot1}\frac{\sigma_{3\cdot1}}{\sigma_{2\cdot1}}$$

$$\therefore \qquad (b_{23\cdot1})\,(b_{32\cdot1}) = r_{23\cdot1}^2$$

$$\therefore \qquad r_{23\cdot1}^2 = \left[-\frac{\sigma_2}{\sigma_3}\frac{\omega_{23}}{\omega_{22}}\right]\left[-\frac{\sigma_3}{\sigma_2}\frac{\omega_{32}}{\omega_{33}}\right]$$

$$= \frac{\omega_{23}^2}{\omega_{22}\omega_{33}}$$

$$\therefore \qquad r_{23\cdot1} = \frac{-\omega_{23}}{\sqrt{\omega_{22}\omega_{33}}} \qquad (\because\ b_{23\cdot1},\ b_{32\cdot1},\ \text{both are with} - \text{ve sign})$$

Now $$\omega_{23} = \begin{vmatrix} r_{11} & r_{12} \\ r_{31} & r_{32} \end{vmatrix}$$

$$= (r_{31}\,r_{12} - r_{32}) \qquad (\because\ r_{11} = 1)$$

$$\omega_{22} = 1 - r_{13}^2,\ \omega_{33} = 1 - r_{12}^2$$

$$\therefore \qquad r_{23\cdot1} = \frac{r_{32} - r_{31}\,r_{12}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{12}^2\right)}}$$

Similarly prove (*ii*)

Ex. 14-6. *In a trivariate distribution*

	$\sigma_1 = 3$	$\sigma_2 = 4,$	$\sigma_3 = 5$
	$r_{12} = 0{\cdot}7$	$r_{13} = 0{\cdot}61,$	$r_{23} = 0{\cdot}4$
find	$r_{23\cdot1},$	$b_{12\cdot3}$ and	$\sigma_{1\cdot23}.$

Sol. We have
$$\omega = \begin{vmatrix} 1 & 0\cdot7 & 0\cdot61 \\ 0\cdot7 & 1 & 0\cdot4 \\ 0\cdot61 & 0\cdot4 & 1 \end{vmatrix} = 0\cdot32$$

$$\omega_{11} = \begin{vmatrix} 1 & 0\cdot4 \\ 0\cdot4 & 1 \end{vmatrix} = 0\cdot84$$

$$\omega_{22} = \begin{vmatrix} 1 & 0\cdot61 \\ 0\cdot61 & 1 \end{vmatrix} = 0\cdot63$$

$$\omega_{33} = \begin{vmatrix} 1 & 0\cdot7 \\ 0\cdot7 & 1 \end{vmatrix} = 0\cdot51$$

$$\omega_{23} = -\begin{vmatrix} 1 & 0\cdot7 \\ 0\cdot61 & 0\cdot4 \end{vmatrix} = 0\cdot027$$

$$\omega_{12} = -\begin{vmatrix} 0\cdot7 & 0\cdot4 \\ 0\cdot61 & 1 \end{vmatrix} = -\ 0\cdot46$$

$$\therefore \quad r_{23\cdot1} = \frac{-\omega_{23}}{\sqrt{\omega_{22}\,\omega_{33}}} = \frac{-\cdot027}{\sqrt{(0\cdot63)(0\cdot51)}}$$

$$= -0\cdot05$$

$$b_{12\cdot3} = -\frac{\sigma_1}{\sigma_2}\cdot\frac{\omega_{12}}{\omega_{11}} = \frac{3}{4}\cdot\frac{0\cdot46}{0\cdot84}$$

$$= 0\cdot54$$

$$\sigma_{1\cdot23} = \sigma_1\sqrt{\frac{\omega}{\omega_{11}}} = 3\sqrt{\frac{0\cdot32}{0\cdot84}} = 1\cdot85.$$

Ex. 14-7. *Show that* $1 - R_{1\cdot23}^2 = \left(1 - r_{12}^2\right)\left(1 - r_{13\cdot2}^2\right)$

Deduce that $R_{1\cdot23} \geq r_{12}$

and $\quad 1 + 2r_{12}\, r_{13}\, r_{23} \geq r_{12}^{\,2} + r_{13}^{\,2} + r_{23}^{\,2}.$

Sol. We have

$$1 - R_{1\cdot23}^2 = \frac{\omega}{\omega_{11}}$$

$$1 - r_{13\cdot2}^2 = 1 - \frac{\omega_{13}^2}{\omega_{11}\omega_{33}}$$

$$= \frac{\omega_{11}\omega_{33} - \omega_{13}^2}{\omega_{11}\omega_{33}}$$

$$= \frac{1}{\omega_{11}\omega_{33}}\left\{\left(1 - r_{23}^2\right)\left(1 - r_{12}^2\right) - \left(r_{21}\, r_{32} - r_{31}\right)^2\right\}$$

$$= \frac{1}{\omega_{11}\omega_{33}}\left\{1 - r_{23}^2 - r_{12}^2 - r_{31}^2 + 2r_{31}r_{21}r_{32}\right\} \qquad \ldots(1)$$

$$= \frac{\omega}{\omega_{11}\omega_{33}}$$

$$\therefore \quad \frac{1 - R_{1\cdot23}^2}{1 - r_{13\cdot2}^2} = \omega_{33} = 1 - r_{12}^2$$

$$\Rightarrow \quad \left(1 - R_{1\cdot23}^2\right) = \left(1 - r_{12}^2\right)\left(1 - r_{13\cdot2}^2\right)$$

Now $\quad 0 \leq r_{13\cdot2}^2 \leq 1$

$$\Rightarrow \quad 0 \leq 1 - r_{13\cdot2}^2 \leq 1 \qquad \ldots(2)$$

$$\therefore \quad 1 - R_{1\cdot23}^2 \leq 1 - r_{12}^2$$

$$\Rightarrow \quad R_{1\cdot23}^2 \geq r_{12}^2$$

$\Rightarrow \quad R_{1\cdot23} \geq$

Also (1) and (2) $\Rightarrow$

$1 - r_{12}^2 - r_{23}^2 - r_{31}^2 + 2r$

$\Rightarrow \quad 1 + 2r_{12}r_{13}r_{23} \geq r_{12}^2 + r_1$

Ex. 14-8. *Show that three re*

$\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} =$

Sol. The equation of three

$\frac{x_1}{\sigma_1}\omega_{11} + \frac{x_2}{\sigma_2}\omega_{12} +$

$\frac{x_1}{\sigma_1}\omega_{21} + \frac{x_2}{\sigma_2}\omega_{22} +$

$\frac{x_1}{\sigma_1}\omega_{31} + \frac{x_2}{\sigma_2}\omega_{32} +$

Planes (1) and (2) will coinc

$\frac{\omega_{11}}{\omega_{21}} =$

and planes (2) and (3) will coinc

$\frac{\omega_{21}}{\omega_{31}} =$

First two ratios in (4) $\Rightarrow$

$\omega_{11}\,\omega_{22} =$

i.e., $\left(1 - r_{23}^2\right)\left(1 - r_{13}^2\right) =$

i.e., $1 - r_{23}^2 - r_{13}^2 + r_{23}^2 r_{13}^2 =$

i.e., $1 - r_{23}^2 - r_{13}^2 - r_{12}^2 + 2r_1$

$\Rightarrow \quad \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} =$

Similarly other ratios in (4)

Ex. 14-9. *Show that*

$b_{12\cdot3}\, b_{23\cdot1}\, b_{31\cdot2} =$

Sol. R.H.S. $=$

$=$

$=$

$=$

Ex. 14-10. *Show that the c*

opposite to that between $x_{1\cdot3}$ *and*

$$\Rightarrow \qquad R_{1\cdot 23} \geq r_{12}.$$

Also (1) and (2) $\Rightarrow$

$$1 - r_{12}^2 - r_{23}^2 - r_{31}^2 + 2r_{31}r_{21}r_{32} \geq 0$$

$$\Rightarrow \quad 1 + 2r_{12}r_{13}r_{23} \geq r_{12}^2 + r_{13}^2 + r_{23}^2 .$$

Ex. 14-8. *Show that three regression planes coincide iff*

$$\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} = 0$$

Sol. The equation of three planes of regression are

$$\frac{x_1}{\sigma_1}\omega_{11} + \frac{x_2}{\sigma_2}\omega_{12} + \frac{x_3}{\sigma_3}\omega_{13} = 0 \qquad \text{...(1)}$$

$$\frac{x_1}{\sigma_1}\omega_{21} + \frac{x_2}{\sigma_2}\omega_{22} + \frac{x_3}{\sigma_3}\omega_{23} = 0 \qquad \text{...(2)}$$

$$\frac{x_1}{\sigma_1}\omega_{31} + \frac{x_2}{\sigma_2}\omega_{32} + \frac{x_3}{\sigma_3}\omega_{33} = 0 \qquad \text{...(3)}$$

Planes (1) and (2) will coincide iff

$$\frac{\omega_{11}}{\omega_{21}} = \frac{\omega_{12}}{\omega_{22}} = \frac{\omega_{13}}{\omega_{23}} \qquad \text{...(4)}$$

and planes (2) and (3) will coincide iff

$$\frac{\omega_{21}}{\omega_{31}} = \frac{\omega_{22}}{\omega_{32}} = \frac{\omega_{23}}{\omega_{32}} \qquad \text{...(5)}$$

First two ratios in (4) $\Rightarrow$

$$\omega_{11}\,\omega_{22} = \omega_{21}\,\omega_{12}$$

i.e.,
$$\left(1 - r_{23}^2\right)\left(1 - r_{13}^2\right) = (r_{32}r_{13} - r_{12})^2$$

i.e.,
$$1 - r_{23}^2 - r_{13}^2 + r_{23}^2 r_{13}^2 = r_{23}^2 r_{13}^2 + r_{12}^2 - 2r_{12}r_{32}r_{13}$$

i.e.,
$$1 - r_{23}^2 - r_{13}^2 - r_{12}^2 + 2r_{12}r_{32}r_{13} = 0$$

$$\Rightarrow \quad \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} = 0.$$

Similarly other ratios in (4) and (5) also imply this condition.

Ex. 14-9. *Show that*

$$b_{12\cdot 3}\, b_{23\cdot 1}\, b_{31\cdot 2} = r_{12\cdot 3}\, r_{23\cdot 1}\, r_{31\cdot 2}.$$

Sol.
$$\text{R.H.S.} = r_{12\cdot 3}\, r_{23\cdot 1}\, r_{31\cdot 2}$$

$$= \left[\frac{-\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}}\right]\left[\frac{-\omega_{23}}{\sqrt{\omega_{22}\omega_{33}}}\right]\left[\frac{-\omega_{31}}{\sqrt{\omega_{33}\omega_{11}}}\right]$$

$$= \left[-\frac{\sigma_1}{\sigma_2}\frac{\omega_{12}}{\omega_{11}}\right]\left[-\frac{\sigma_2}{\sigma_3}\frac{\omega_{23}}{\omega_{22}}\right]\left[-\frac{\sigma_3}{\sigma_1}\frac{\omega_{31}}{\omega_{33}}\right]$$

$$= b_{12\cdot 3}\, b_{23\cdot 1}\, b_{31\cdot 2}.$$

Ex. 14-10. *Show that the correlation coefficient between $x_{1\cdot 23}$ and $x_{2\cdot 13}$ is equal ana opposite to that between $x_{1\cdot 3}$ and $x_{2\cdot 3}$.*

Sol. We have

$$\text{Cov}\,(x_{1\cdot 23}, x_{2\cdot 13}) = \frac{1}{N}\,\Sigma x_{1\cdot 23}\, x_{2\cdot 13}$$

$$= \frac{1}{N}\,\Sigma x_{1\cdot 23}\,(x_2 - b_{21\cdot 3}\, x_1 - b_{23\cdot 1}\, x_3)$$

$$= -\, b_{21\cdot 3}\,\frac{1}{N}\,\Sigma x_{1\cdot 23}\, x_1$$

$$= -\, b_{21\cdot 3}\,\frac{1}{N}\,\Sigma x_{1\cdot 23}{}^2$$

$$= -\, b_{21\cdot 3}\,\sigma_{1\cdot 23}^2$$

$$\therefore \qquad r(x_{1\cdot 23}, x_{2\cdot 13}) = \frac{\text{Cov}\,(x_{1\cdot 23}, x_{2\cdot 13})}{\sigma_{1\cdot 23}\,\sigma_{2\cdot 13}}$$

$$= -\, b_{21\cdot 3}\,\frac{\sigma_{1\cdot 23}}{\sigma_{2\cdot 13}}.$$

$$= -\left[-\frac{\sigma_2}{\sigma_1}\,\frac{\omega_{21}}{\omega_{22}}\right]\left[\frac{\sigma_1^2}{\sigma_2^2}\,\frac{\omega/\omega_{11}}{\omega/\omega_{22}}\right]^{1/2}$$

$$= -\left\{-\frac{\omega_{21}}{(\omega_{11}\,\omega_{22})^{1/2}}\right\}$$

$$= -\, r_{21\cdot 3}$$

$$= -\, r(x_{1\cdot 3}, x_{2\cdot 3}).$$

Ex. 14-11. *Show that if* $x_3 = ax_1 + bx_2$, the three partial correlations are numerically equal to unity, $r_{13\cdot 2}$ *having the sign of* a, $r_{23\cdot 1}$, *the sign of* b *and* $r_{12\cdot 3}$ *the opposite sign of* $\frac{a}{b}$.

Sol. Here x_1, x_2 can be regarded as independent and x_3 is dependent on both of them

$$\therefore \qquad r_{12} = 0 \Rightarrow \text{Cov}\,(x_1, x_2) = 0$$

$$\Rightarrow \Sigma x_1 x_2 = 0$$

Now $\qquad \text{Var}\,(x_3) = \text{Var}\;(ax_1 + bx_2)$

$$\therefore \qquad \sigma_3{}^2 = a^2\,\text{Var}\,(x_1) + b^2\,\text{Var}\,(x_2)$$

$$= a^2\,\sigma_1{}^2 + b^2\,\sigma_2{}^2$$

$$\text{Cov}\,(x_1, x_3) = \frac{1}{N}\,\Sigma x_1\, x_3$$

$$= \frac{1}{N}\,\Sigma x_1\,(ax_1 + bx_2) = a\sigma_1{}^2$$

$$\therefore \qquad r_{13} = \frac{a\,\sigma_1^2}{\sigma_1\sigma_3} = a\,\frac{\sigma_1}{\sigma_3}$$

Similarly $\qquad r_{23} = b\,\dfrac{\sigma_2}{\sigma_3}$

$$\therefore \qquad r_{13\cdot 2} = \frac{r_{13} - r_{12}\, r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}}$$

$$= \frac{a\sigma_1/\sigma_3}{\sqrt{1-\dfrac{b^2\sigma_2^2}{\sigma_3^2}}} = \frac{a\sigma_1}{\sqrt{\sigma_3^2 - b^2\,\sigma_2^2}}$$

=

=

Similarly $\qquad r_{23\cdot 1} =$

Now $\qquad r_{12\cdot 3} =$

=

=

=

according as $ab >$ or < 0 *i.e.*, a an

i.e., $\qquad \dfrac{a}{b}$

$\therefore \quad r_{12\cdot 3}$ has sign opposite o

Ex. 14-12. *If* $r_{23} = 1$, *show th*

$$R_{1\cdot 23}^2 =$$

and $\qquad \sigma_{1\cdot 23}^2 =$

Sol. We have $R_{1\cdot 23}^2\,(1 - r_{23}^2)$

Put $\qquad r_{23} =$

$$(r_{12} - r_{13})^2 = ($$

$$\therefore \qquad R_{1\cdot 23}^2\,(1 - r_{23}^2) =$$

$$\Rightarrow R_{1\cdot 23}^2 \to \left[\frac{2r_{12}^2}{1+r_{23}}\right]_{r_{23}=1} =$$

By Ex. 14-3,

$$\sigma_{1\cdot 23}^2 = ($$

Ex. 14-13. *Show that* $R_{1\cdot 23}^2 =$

Sol. $\qquad$ R.H.S. $= b$

=

= −

$$= \frac{a\sigma_1}{\sqrt{a^2\sigma_1^2 + b^2\sigma_2^2 - b^2\sigma_2^2}} = \frac{a\sigma_1}{|a|\sigma_1}$$

$$= \pm 1 \text{ according as } a > \text{ or } < 0.$$

Similarly $\quad r_{23\cdot1} = \pm 1$ according as $b >$ or < 0.

Now
$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$= \frac{-\dfrac{a\sigma_1}{\sigma_3}\cdot\dfrac{b\sigma_2}{\sigma_3}}{\sqrt{\left[1 - \dfrac{a^2\sigma_1^2}{\sigma_3^2}\right]\left[1 - \dfrac{b^2\sigma_2^2}{\sigma_3^2}\right]}}$$

$$= \frac{-ab\sigma_1\sigma_2}{\sqrt{(\sigma_3^2 - a^2\sigma_1^2)(\sigma_3^2 - b^2\sigma_2^2)}}$$

$$= \frac{-ab}{|ab|} = \mp 1$$

according as $ab >$ or < 0 *i.e.*, a and b are of same or opposite signs

i.e., $\quad \dfrac{a}{b} >$ or < 0

$\therefore \quad r_{12\cdot3}$ **has sign opposite of** $\dfrac{a}{b}$.

Ex. 14-12. *If* $r_{23} = 1$, *show that*

$$R_{1\cdot23}^2 = r_{12}^2 = r_{13}^2$$

and
$$\sigma_{1\cdot23}^2 = \sigma_1^2(1 - r_{12}^2).$$

Sol. We have $R_{1\cdot23}^2(1 - r_{23}^2) = r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}$

Put
$$r_{23} = 1$$
$$(r_{12} - r_{13})^2 = 0 \Rightarrow r_{12} = r_{13}$$

$$\therefore \quad R_{1\cdot23}^2(1 - r_{23}^2) = 2r_{12}^2(1 - r_{23})$$

$$\Rightarrow R_{1\cdot23}^2 \to \left[\frac{2r_{12}^2}{1 + r_{23}}\right]_{r_{23}=1} = r_{12}^2 = r_{13}^2$$

By Ex. 14-3,

$$\sigma_{1\cdot23}^2 = \sigma_1^2(1 - R_{1\cdot23}^2) = \sigma_1^2(1 - r_{12}^2).$$

Ex. 14-13. *Show that* $R_{1\cdot23}^2 = b_{12\cdot3}\,r_{12}\dfrac{\sigma_2}{\sigma_1} + b_{13\cdot2}\,r_{13}\dfrac{\sigma_3}{\sigma_1}$.

Sol.
$$\text{R.H.S.} = b_{12\cdot3}\,r_{12}\frac{\sigma_2}{\sigma_1} + b_{13\cdot2}\,r_{13}\frac{\sigma_3}{\sigma_1}$$

$$= \left[-\frac{\sigma_1}{\sigma_2}\frac{\omega_{12}}{\omega_{11}}\right]r_{12}\frac{\sigma_2}{\sigma_1} + \left[-\frac{\sigma_1}{\sigma_3}\frac{\omega_{13}}{\omega_{11}}\right]r_{13}\frac{\sigma_3}{\sigma_1}$$

$$= -\frac{1}{\omega_{11}}\{\omega_{12}r_{12} + \omega_{13}r_{13}\}$$

$$= -\frac{1}{1-r_{23}^2}\{r_{12}(r_{31}r_{23}-r_{21})+r_{13}(r_{21}r_{32}-r_{31})\}$$

$$= \frac{1}{1-r_{23}^2}\left\{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{32}\right\}$$

$$= R_{1\cdot23}^2.$$

Ex. 14-14. *Show that* $R_{1\cdot23}=0 \Rightarrow r_{12}=r_{13}=0.$

Sol. We have $1-R_{1\cdot23}^2 = \left(1-r_{12}^2\right)\left(1-r_{13\cdot2}^2\right)$ (See Ex. 14-7)

Put $R_{1\cdot23}^2 = 0$

$$\therefore \quad \left(1-r_{12}^2\right)\left(1-r_{13\cdot2}^2\right) = 1 \qquad ...(1)$$

Since $0 \le r_{12}^2 \le 1$ and $0 \le r_{13\cdot2}^2 \le 1$,

(1) is possible only when

$$1-r_{12}^2 = 1 \Rightarrow r_{12}=0 \qquad ...(2)$$

and

$$1-r_{13\cdot2}^2 = 1 \Rightarrow r_{13\cdot2}=0 \qquad ...(3)$$

$$(3) \Rightarrow \quad r_{13}-r_{12}r_{32} = 0 \qquad \text{(see Ex. 14-5)}$$

$$\Rightarrow \quad r_{13} = 0 \text{ (using 2).}$$

EXERCISES

1. Find the regression equation of x_3 on x_1 and x_2 given that

$r_{12}=0\cdot28$ $\quad r_{23}=0\cdot49$ $\quad r_{31}=0\cdot51$

$\sigma_1=2\cdot7$ $\quad \sigma_2=2\cdot4$ $\quad \sigma_3=2\cdot7$

Also find $R_{1\cdot23}$, $r_{23\cdot1}$.

2. Calculate the multiple correlation coefficient of x_1 on x_2 and x_3 from the following data :

x_1	1	2	3	4	5	8
x_2	2	2	4	2	2	4
x_3	13	15	21	17	21	32

Also find the regression equation of x_1 on x_2, x_3.

3. Let x_1 = seed-hay crop, x_2 = rainfall and x_3 = accumulated temperature. The following means, s.ds. and correlations are found

$\bar{x}_1=28\cdot02,$ $\quad \bar{x}_2=4\cdot9,$ $\quad \bar{x}_3=594$

$\sigma_1=4\cdot4$ $\quad \sigma_2=1\cdot1,$ $\quad \sigma_3=85$

$x_{12}=0\cdot8,$ $\quad r_{13}=-0\cdot4,$ $\quad r_{23}=-0\cdot56.$

Find all partial correlations and the regression equations for hay-crop on rainfall and accumulated temperature.

4. Let x_1, x_2, x_3 are variates with zero means that

$\sigma_1=1,$ $\quad \sigma_2=1\cdot3,$ $\quad \sigma_3=1\cdot9$

$r_{12}=\cdot37,$ $\quad r_{13}=-0\cdot641,$ $\quad r_{23}=-0\cdot736.$

Verify that $\quad r_{13\cdot2}=r_{43\cdot2}$ $\quad$ where $x_4=x_1+x_2$.

5. If $x_1=y_1+y_2, x_2=y_2+y_3, x_3=y_3+y_1$ where y_1, y_2, y_3 are uncorrelated variables each of which has zero mean and unit standard diviation, find $R_{1\cdot23}$. $\left(\textbf{Ans. } \frac{1}{\sqrt{3}}\right)$

6. If $r_{12}=r_{13}=r_{23}=\rho\ (\neq -1)$, show that each partial correlation coefficient is $\frac{\rho}{1+\rho}$ and each multiple correlation coefficient is

$$\frac{\rho\sqrt{2}}{\sqrt{1+\rho}}$$

Also show that $1-R_{1\cdot23}^2 =$ ·

7. x_1, x_2, x_3 are uncorrelated var

Let $\quad y_1 =$

show that y_1, y_2, y_3 are standa

8. If $a_1x_1+a_2x_2+a_3x_3=k$, pro

$r_{12} =$ -

with two similar expressions

Also show that all the partial q

9. If x_1, x_2, x_3 are three variates

expected value of x_1 for given

and x_3, prove that

Cov $(x_1, e_1) =$

10. If x_1, x_2, x_3 are standard varia

$E(x_2x_3) =$

11. If $\quad r_{23} =$ (

$R_{1\cdot23}^2 =$

and $\quad \sigma_{1\cdot23}^2 =$ 1

12. Suppose a computer has foun

$r_{12}=0\cdot6,$

Examine whether his comput

(**Hint.** Find $r_{12\cdot3}$).

13. Comment on the consistancy

$r_{12}=0\cdot6,$

14. For what value of $R_{1\cdot23}$ will x

15. If r_{12} and r_{13} are given, show

$r_{12}r_{13} \pm$ (

(**Hint.** Use $r_{12\cdot3}^2 \le 1$).

16. If $r_{12}=k$; $r_{23}=-k$, show that

17. A number of persons are mea

and product moment correlat

r_{12}

[**Hint.** Use $E\left[\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}+\right.$

18. Show that $\quad b_{12.3} =$

Also show that $1 - R_{1\cdot 23}^2 = \dfrac{(1-\rho)(1+2\rho)}{1+\rho}$.

7. x_1, x_2, x_3 are uncorrelated variates with same variance.

Let
$$y_1 = \frac{x_1 - x_3}{\sqrt{2}},\ y_2 = \frac{x_1 + x_2 + x_3}{\sqrt{3}},\ y_3 = \frac{x_1 + 2x_2 + x_3}{\sqrt{6}}$$
show that y_1, y_2, y_3 are standard variates. Also find $r_{12\cdot 3}$ and $R_{1\cdot 23}$ for y's.

8. If $a_1x_1 + a_2x_2 + a_3x_3 = k$, prove that
$$r_{12} = \frac{\left(a_3^2\sigma_3^2 - a_1^2\sigma_1^2 - a_2^2\sigma_2^2\right)}{2a_1a_2\sigma_2}$$
with two similar expressions for r_{13} and r_{23}.
Also show that all the partial quotients are equal to -1 provided that a's are all positive.

9. If x_1, x_2, x_3 are three variates measured from their respective means and if e_1 is the expected value of x_1 for given values of x_2 and x_3 from the linear regression of x_1 on x_2 and x_3, prove that
$$\text{Cov}(x_1, e_1) = \text{Var}(e_1) = \text{Var}(x_1) - \text{Var}(x_1 - e_1).$$

10. If x_1, x_2, x_3 are standard variates and
$$E(x_2x_3) = E(x_1x_3) = \frac{1}{2}.\ \text{Show that}$$
$$E(x_1x_2) \geq -5/2.$$

11. If $r_{23} = 0$, show that
$$R_{1\cdot 23}^2 = r_{12}^2 + r_{13}^2$$
and
$$\sigma_{1\cdot 23}^2 = 1 - r_{12}^2 - r_{13}^2.$$

12. Suppose a computer has found, for a given set of values of x_1, x_2 and x_3
$r_{12} = 0{\cdot}6$, $r_{23} = 0{\cdot}7$, $r_{31} = -0{\cdot}4$.
Examine whether his computations may be said to be free from error. (**Ans.** No)
(**Hint.** Find $r_{12\cdot 3}$).

13. Comment on the consistancy of
$r_{12} = 0{\cdot}6$, $r_{23} = 0{\cdot}8$, $r_{31} = -0{\cdot}5$.

14. For what value of $R_{1\cdot 23}$ will x_2 and x_3 be uncorrelated ?

15. If r_{12} and r_{13} are given, show that r_{23} must lie in the range
$$r_{12}r_{13} \pm \left(1 - r_{12}^2 - r_{13}^2 + r_{12}^2 r_{13}^2\right)^{1/2}$$
(**Hint.** Use $r_{12\cdot 3}^2 \leq 1$).

16. If $r_{12} = k$; $r_{23} = -k$, show that r_{13} will lie between -1 and $1 - 2k^2$.

17. A number of persons are measured for heights x_1, weights x_2 and chest expansions x_3 and product moment correlation co-efficients are calculated. Show that
$$r_{12} + r_{23} + r_{31} \geq -3/2.$$
$$\left[\textbf{Hint.}\quad \text{Use } E\left[\frac{x_1}{\sigma_1} + \frac{x_2}{\sigma_2} + \frac{x_3}{\sigma_3}\right]^2 \geq 0\right]$$

18. Show that
$$b_{12.3} = \frac{b_{12} - b_{13}b_{32}}{1 - b_{23}b_{32}}.$$

15

Sampling Theory and Large Sample Tests

15.1. Introduction

Very often in practice one is interested in drawing valid conclusions about a large group of individuals or objects. Instead of examining the entire group (which may be difficult or impossible) one may think of examining a small part of it. This is done with the aim of inferring certain facts about the large group from the results found for smaller part. This process is called **statistical inference**. Various technical words used during this process are explained as below :

Population. *Any collection of individuals or of attributes or the results of operations which can be specified numerically.*

Finite Population. *Population containing finite number of members. Otherwise the population is called infinite population e.g., the population of boys in a college and the population of pressures at various points in the atmosphere are finite and infinite respectively.*

Existent Population. *Population of concrete objects e.g., the population of a city.*

Hypothetical Population. *Population of non-concrete objects e.g., the population of heads and tails obtained by tossing a coin an infinite number of times.*

Sample. *A part or small section selected from the population is called a* **sample** *and the process of such selection is called* **sampling.**

Random Sampling. *When a sample is taken in such a way that each member of the population has the same chance of being selected, the sample obtained is called* **random sample** *and the technique is called* **random sampling.**

Simple Sampling. *When a random sample is drawn from a population in such a way that the chance of selection of a member at any stage is independent of previous selections, the sample obtained is called* **simple sample** *and the technique is called* **simple sampling.**

Stratified Sampling. *In this process the entire heterogeneous population is divided into a number of homogeneous groups (termed as strata) which differ from one another but each of these is homogeneous within itself. The samples are drawn from each stratum (the sample size in each stratum varying according to the relative importance of the stratum in the population). The aggregate of the samples from each of the stratum is called* **stratified sample** *and the technique is called* **stratified sampling** *e.g. To estimate the average income of the inhabitants of a city, it is necessary that all sections of the society must be included in the sample otherwise there is a likelihood that more rich people or poor people many be dominating the sample. For this purpose, it is better to divide the city into different strata say, according to the localities; slums, middle-class localities and bungalow areas and then to draw samples from each of these localities. This would ensure that all sections of the society are represented in the sample.*

Sampling with or without Replacei

Sampling where each number of
sampling with replacement and if
called **sampling without replaceme**

Remark. (*i*) From a finite popu
drawn without exhausting the popula

(*ii*) For most practical purposes s
can be considered as sampling from

Parameters

A population is considered to be
$f(x)$ of the associated variable x is knc
said to be normal.

Certain quantities may appear ir
Other quantities such as mean, variar
quantities are called **population para**

Remark. When the population
known.

Statistic. It is a statistical measu

Remark. Statistic is calculated w
To each population parameter there
statistic may not always give the best
theory is to decide how to form a pro
given population parameter.

Sampling Distribution

The statistic is itself a random vari
distribution. It can be thought of as b

All possible samples of given siz
the statistic is calculated. The values

Standard Errors. *The standard*
known as standard error and is writte

Precision. *The reciprocal of S.E.*

Probable Error (P.E.). *It is defi*

$P.E. = (0\cdot$

Standard Errors of Various Parame

(*i*) *Quartiles*

(*ii*) *Median*

(*iii*) *S.D.*

(*iv*) *Variance*

(*v*) *Co-efficient of correlation*

Sampling with or without Replacement

Sampling where each number of a population may be chosen more than once is called **sampling with replacement** and if each member cannot be chosen more than once, it is called **sampling without replacement.**

Remark. (*i*) From a finite population, a sample with replacement of any size can be drawn without exhausting the population.

(*ii*) For most practical purposes sampling from a finite population (which is very large) can be considered as sampling from an infinite population.

Parameters

A population is considered to be known, if the probability function (or density function) $f(x)$ of the associated variable x is known *e.g.*, if x is normally distributed, the population is said to be normal.

Certain quantities may appear in $f(x)$ (*e.g.*, m and σ in case of normal distribution). Other quantities such as mean, variance etc., can then be obtained in terms of these. Such quantities are called **population parameters** or **simply parameters.**

Remark. When the population is given, the population parameters are taken to be known.

Statistic. It is a statistical measure computed from sample observations alone.

Remark. Statistic is calculated with the purpose of estimating a population parameter. To each population parameter there is a statistic to be computed from the sample. This statistic may not always give the best estimate. Once of the important problems of sampling theory is to decide how to form a proper sample statistic, so as to get a best estimate of a given population parameter.

Sampling Distribution

The statistic is itself a random variate. Its probability distribution is often called **sampling distribution.** It can be thought of as below :

All possible samples of given size are taken from the population and for each sample the statistic is calculated. The values of the statistic form its sampling distribution.

Standard Errors. *The standard deviation of a sampling distribution of a statistic is known as standard error and is written as 'S.E.'.*

Precision. *The reciprocal of S.E. is called precision.*

Probable Error (P.E.). *It is defined by*

$$P.E. = (0{\cdot}67449)\, S.E.$$

Standard Errors of Various Parameters

(*i*) *Quartiles*	$1{\cdot}36263 \dfrac{\sigma}{\sqrt{n}}$.
(*ii*) *Median*	$1{\cdot}25331 \dfrac{\sigma}{\sqrt{n}}$.
(*iii*) *S.D.*	$\dfrac{\sigma}{\sqrt{2n}}$.
(*iv*) *Variance*	$\sigma^2 \sqrt{\dfrac{2}{n}}$.
(*v*) *Co-efficient of correlation*	$\dfrac{(1-r^2)}{\sqrt{n}}$.

(*vi*) μ_3 $\qquad\qquad$ $\sigma^3\sqrt{\dfrac{6}{n}}$.

Unbiased Estimate. *A statistic 't' is said to be an unbiased estimate of a parameter* θ *if* $E(t) = \theta$.

Asymptotically Unbiased Estimate. *A statistic 't_n' is said to be an asymptotically unbiased estimate of a parameter* θ *if*

$$\underset{n\to\infty}{\text{Lt}}\ E(t_n) = \theta$$

where n is the size of the sample.

Large and Small Samples. *Samples of size greater than* 30 *are called large samples and of size less than or equal to* 30 *are called small samples.*

Hypothesis. *Very often it is required to make decisions about populations on the basis of sample information. Such decisions are called* **Statistical decisions.** *In attempting to reach decisions, it is often necessary to make assumptions about the population involved. Such assumptions, which are not necessarily true, are called statistical hypothesis.*

Null Hypothesis. *The hypothesis tested for possible rejection under the assumption that it is true is usually called null hypothesis.*

Tests of Significance. *Procedures which enable us to decide, on the basis of sample information, whether to accept or reject hypothesis or to determine whether observed sampling results differ significantly from expected results are called tests of significance, rules of decision or tests of hypothesis.*

Level of significance. *The probability level below which we reject the hypothesis is called the level of significance.*

Confidence Interval

It is the interval in which a population parameter is expected to lie with certain probability (mentioned in percentage).

The end numbers are called **confidence limits** or **fiducial limits.** The probability is called **confidence level.**

15.2. Sampling of Attributes

In the case of sampling of attributes we are concerned only with the presence or absence of some given attribute. The selection of an individual in sampling may be called a trial and the presence of a specified attribute a success and its absence a failure.

By simple sampling of attributes we mean random sampling in which each trial has the same chance of success and in which the chances of success of different trials are independent whether the previous trials have been made or not.

Mean and s.d.

Suppose we are to draw a simple sample of n individuals from a population. Let p be the chance of success and q the chance of failure of each trial.

Then $\qquad p + q = 1$

The drawing of a sample is identical with the problem of a series of n independent trials with constant probability p of success.

$\therefore$ The probability of 0, 1, 2, ..., n successes are the successive terms in the binomial expansion of $(q + p)^n$. (from B.D.)

$\therefore$ The probability of x successes is given by

$$P(x) = {}^nc_x\, p^x\, q^{n-x},\ x = 0, 1, 2, ..., n.$$

The binomial probability distribution so obtained is called sampling distribution of the number of successes in the sample.

$\therefore$ The expected value or mea

i.e., $\qquad E(x) =$

and the s.d. of the number of succ

$=$

Now proportion of successes =

$\therefore \qquad E\left(\dfrac{x}{n}\right) =$

and $\qquad \text{Var}\left(\dfrac{x}{n}\right) =$

$\therefore$ Standard deviation of $\dfrac{x}{n} =$

15.2.1. To test the significance of

Let us suppose that *w.r.t.* attribu into two mutually exclusive and individuals possessing A in a singl possessing A is given by

$p =$

Let P be the probability for a success.

Then x is a binomial variate wit

$\therefore \qquad u =$

is a binomial variate with mean zero a

$\therefore \qquad P(|u| > 3) = 1$

$= 1$

Similarly $\quad P(|u| > 1{\cdot}96) = 0{\cdot}$

$p(|u| > 2{\cdot}58) = 0{\cdot}$

The probability P is obtained b obtained above the rules for taking d

(*i*) If $|u| > 3$, the difference bet is highly significant and hence the h

(*ii*) If $2{\cdot}58 < |u| < 3$, the differe

(*iii*) If $1{\cdot}96 < |u| < 2{\cdot}58$, the diff

(*iv*) If $|u| < 1{\cdot}96$, the difference with the hypothesis and hence the divergence may be due to flactuation

Note. (*i*) The above test is valid distribution may not be nearly norma

(*ii*) The test may furnish evid hypothesis to be correct. It can at the

(*iii*) Since the hypothesis can be is set, *e.g.*, to test whether there is an

$\therefore$ The expected value or mean value of the number of successes

i.e., $$E(x) = np$$

and the s.d. of the number of successes

$$= \sqrt{npq}.$$

Now proportion of successes $= \dfrac{x}{n}$

$$\therefore \quad E\left(\frac{x}{n}\right) = \frac{1}{n} \cdot E(x) = p$$

and $$\text{Var}\left(\frac{x}{n}\right) = \frac{1}{n^2}\text{Var}(x) = \frac{pq}{n}$$

$\therefore$ Standard deviation of $\dfrac{x}{n} = \sqrt{\dfrac{pq}{n}}$.

15.2.1. To test the significance of single proportion for large samples

Let us suppose that *w.r.t.* attribute A, it is possible to classify individuals of a population into two mutually exclusive and collectively exhaustive sets. Let x be the number of individuals possessing A in a single sample of size n. Then the proportion of individuals possessing A is given by

$$p = \frac{x}{n}.$$

Let P be the probability for an individual to possess A *i.e.,* P is the probability of success.

Then x is a binomial variate with expected value nP and s.d. $\sqrt{nPQ}$. (where $Q = 1 - P$)

$$\therefore \quad u = \frac{x - nP}{\sqrt{nPQ}} = \frac{p - P}{\sqrt{\dfrac{PQ}{n}}}$$

is a binomial variate with mean zero and *s.d.* unity. Since sample is large u is nearly a $N(0, 1)$.

$$\therefore \quad P(|u| > 3) = 1 - p(-3 < u < 3) = 1 - 2\,P(0 < u < 3)$$
$$= 1 - 2(0{\cdot}49865) = 1 - 0{\cdot}9973 = 0{\cdot}0027$$

Similarly $$P(|u| > 1{\cdot}96) = 0{\cdot}05$$
$$p(|u| > 2{\cdot}58) = 0{\cdot}01.$$

The probability P is obtained by setting null hypothesis. On the basis of probabilities obtained above the rules for taking decisions are :

(*i*) If $|u| > 3$, the difference between the observed and expected number of successes is highly significant and hence the hypothesis is certainly wrong and is to be rejected.

(*ii*) If $2{\cdot}58 < |u| < 3$, the difference is significant at 1% level of significance.

(*iii*) If $1{\cdot}96 < |u| < 2{\cdot}58$, the difference is significant at 5% level of significance.

(*iv*) If $|u| < 1{\cdot}96$, the difference is not significant and the data is said to be consistent with the hypothesis and hence the hypothesis may be accepted. We can also say that divergence may be due to flactuations of sampling —

Note. (*i*) The above test is valid only for large samples since for small samples binomial distribution may not be nearly normal.

(*ii*) The test may furnish evidence against the hypothesis but it cannot prove the hypothesis to be correct. It can at the most provide no evidence against it.

(*iii*) Since the hypothesis can be rejected but cannot be proved, always null hypothesis is set, *e.g.*, to test whether there is any difference, it is assumed that there is no difference :

to test whether there is any relationship, it is assumed that there is no relationship etc. The rejection of no difference will mean a difference and the rejection of no relationship a relationship.

Ex. 15-1. *A coin is tossed* 400 *times and it turns up head* 216 *times. Discuss whether the coin may be regarded as unbiased one.*

Sol. Let x be the number of heads obtained and P the prob. of getting head in a toss. Set the hypothesis : 'Coin is unbiased'.

Then $\qquad P = \dfrac{1}{2}.$

Here $\qquad n = 400,$ and $x = 216$

$$\therefore \qquad u = \frac{x-nP}{\sqrt{nPQ}} = \frac{216-400\cdot\frac{1}{2}}{\sqrt{\frac{1}{2}\cdot\frac{1}{2}\cdot 400}} = 1{\cdot}6 < 1{\cdot}96$$

$\therefore$ The hypothesis may be correct and hence the coin may be regarded as unbiased.

Ex. 15-2. *In some dice-throwing experiment, Weldon threw dice* 49,152 *times and of these* 25,145 *yielded a,* 4, 5 *or* 6. *Is this consistent with the hypothesis that the dice were unbiased ?*

Sol. Set the hypothesis 'Dice was unbiased'

Then P = prob. of 4, 5 or 6 in a throw $= \dfrac{3}{6} = \dfrac{1}{2}$

Here $n = 49{,}152$ and $x = 25{,}145$

$$\therefore \qquad u = \frac{25145-49152\frac{1}{2}}{\sqrt{49152\,\frac{1}{2}\cdot\frac{1}{2}}} \simeq \frac{569}{110{\cdot}85} \simeq 5{\cdot}13 > 3.$$

$\therefore$ Hypothesis is wrong and hence the dice could not be regarded as unbiased.

Ex. 15-3. *Certain crosses of the pea gave* 5,321 *yellow and* 1,804 *green seeds. The expectation is* 25% *green seeds on a Mendelian hypothesis. Is the divergence significant or might have occurred as due to fluctuations of simple sampling ?*

Sol. Total number of seeds (n) = 5321 + 1804 = 7125

Here P = expected proportions of green seeds

$$= \frac{25}{100} = \frac{1}{4}$$

$\therefore$ The standard error of green seeds

$$= \sqrt{7125\cdot\frac{1}{4}\cdot\frac{3}{4}} \simeq 36{\cdot}6$$

$$\therefore \qquad u = \frac{1804-\frac{1}{4}\cdot 7125}{36{\cdot}6} = 0{\cdot}6 < 1{\cdot}96$$

$\therefore$ The data is consistent with the hypothesis and hence the divergence may be regarded as due to fluctuations of simple sampling.

Ex. 15-4. *A die is thrown* 9,000 *times and a throw of* 3 *or* 4 *is reckoned as a success. Suppose that* 3,240 *throws of a* 3 *or* 4 *have been made out. Do the data indicate an unbiased die ? If not, find the probable limits of prob. of getting* 3 *or* 4.

Sol. Here $\qquad n = 9{,}000$

$\qquad x = 3{,}240$

Set the hypothesis : 'Die is unb

Then $P = $

$\therefore \quad u = $

$\therefore$ Difference is highly signific

$\therefore$ The die cannot be regarded

$\therefore \quad P \neq \dfrac{1}{3}$

In order to find the limits of P successes from the sample.

Now proportion of successes

$= \dfrac{3}{9}$

$\therefore$ Estimate of the standard erro

$\left|\dfrac{\frac{x}{n}-P}{0{\cdot}005}\right| < 3$ *i.e.*, $\dfrac{x}{n} - 3(0\cdot$

$\therefore$ Probable limits of P are giv

$\therefore$ Probable limits of P are

$0{\cdot}36 \mp 3(0{\cdot}005)$ *i.e.*, $0{\cdot}34$

Ex. 15-5. *In a locality of* 18,00 *these* 840 *families,* 206 *families were desired to estimate how many out of t less. Within what limits would you pl*

Sol. Let p be the proportion of f

Then estimate of p from the sam

$= \dfrac{20}{84}$

$\therefore$ Estimate of the standard error

$\therefore$ Probable limits of p are

$0{\cdot}245 \mp 3(0{\cdot}015)$ *i.e.*, $0{\cdot}20$

$\therefore$ The probable limits of the nu and 5220.

Ex. 15-6. *A sample of* 900 *day district and* 100 *of them are found to b of foggy days in the district?*

Set the hypothesis : 'Die is unbiased'.

Then $$P = \frac{2}{6} = \frac{1}{3}$$

$$\therefore \quad u = \frac{3240 - 9000 \cdot \frac{1}{3}}{\sqrt{9000 \cdot \frac{1}{3} \cdot \frac{2}{3}}} \simeq 5{\cdot}4 > 3.$$

$\therefore$ Difference is highly significant and hence the hypothesis is wrong.
$\therefore$ The die cannot be regarded as unbiased.

$$\therefore \quad P \neq \frac{1}{3}$$

In order to find the limits of P, we estimate the standard error of the proportion of successes from the sample.

Now proportion of successes

$$= \frac{3240}{9000} = 0{\cdot}36$$

$\therefore$ Estimate of the standard error of the proportion of successes

$$= \sqrt{\frac{(0{\cdot}36)\,(1 - 0{\cdot}36)}{9000}} = \sqrt{\frac{(0{\cdot}36)(0{\cdot}64)}{9000}} \simeq 0{\cdot}005$$

$\therefore$ Probable limits of P are given by

$$\left| \frac{\frac{x}{n} - P}{0{\cdot}005} \right| < 3 \text{ i.e., } \frac{x}{n} - 3(0{\cdot}005) < P < \frac{x}{n} + 3(0{\cdot}005)$$

$\therefore$ Probable limits of P are

$0{\cdot}36 \mp 3(0{\cdot}005)$ *i.e.*, $0{\cdot}345$ and $0{\cdot}375$.

Ex. 15-5. *In a locality of 18,000 families a sample of 840 families was selected. Of these 840 families, 206 families were found to have a monthly income of Rs. 50 or less. It is desired to estimate how many out of the 18,000 families have a monthly income of Rs. 50 or less. Within what limits would you place your estimate ?*

Sol. Let p be the proportion of families with income Rs. 50 or less in the locality.
Then estimate of p from the sample

$$= \frac{206}{840} = \frac{103}{420} = 0{\cdot}245$$

$\therefore$ Estimate of the standard error of the proportion of families with income Rs. 50 or less

$$= \sqrt{\frac{103}{420}\left(1 - \frac{103}{420}\right)\frac{1}{840}} \simeq 0{\cdot}015$$

$\therefore$ Probable limits of p are

$0{\cdot}245 \mp 3(0{\cdot}015)$ *i.e.*, $0{\cdot}20$ and $0{\cdot}229$

$\therefore$ The probable limits of the number of families with income Rs. 50 or less are 3600 and 5220.

Ex. 15-6. *A sample of 900 days is taken from meteorological records of a certain district and 100 of them are found to be foggy. What are the probable limits to the percentage of foggy days in the district?*

Sol. Proportion of foggy days in the district (estimated from the sample)

$$= \frac{100}{900} = \frac{1}{9} = 0{\cdot}1111$$

$\therefore$ Estimate of the standard error of the proportion of foggy days in the district

$$= \sqrt{\frac{1}{900}\cdot\frac{1}{9}\left(1-\frac{1}{9}\right)} = 0{\cdot}0105$$

$\therefore$ Probable limits to the percentage of foggy days are

$100\ \{0{\cdot}1111 \mp 3(0{\cdot}0105)\}$ *i.e.*, 7·96% and 14·26%.

Ex. 15-7. *A sample of* 500 *pineapples was taken from a large consignment and* 65 *were found to be bad. Estimate the proportion of bad pineapples in the consignment, as well as the standard error of the estimate. Deduce that the percentage of bad pineapples in the consignment almost certainly lies between* 8·5 *and* 17·5.

Sol. Here $n = 500$,

p = proportion of bad pineapples in a consignment

$$= \frac{65}{500} = 0{\cdot}13.$$

$\therefore$ The standard error of the proportion of bad pineapples

$$= \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0{\cdot}13)(0{\cdot}87)}{500}} \simeq 0{\cdot}015$$

$\therefore$ Probable limits to the percentage of bad pineapples are

$100\ \{0{\cdot}13 \mp 3(0{\cdot}015)\}$ *i.e.*, 8·5% and 17·5%.

Ex. 15-8. *A biased coin was thrown* 400 *times and head resulted* 240 *times. Find the standard error of the observed proportion of heads and deduce that the probability of getting a head in a throw of the coin lies almost certain between* 0·53 *and* 0·67.

Sol. Observed proportion of heads

$$= \frac{240}{400} = 0{\cdot}6$$

$\therefore$ S.E. of the observed proportion of heads

$$= \sqrt{\frac{(0{\cdot}6)(0{\cdot}4)}{400}}$$

$$= 0{\cdot}0245$$

$\therefore$ The probability of getting a head in a throw of a coin lies in

$$0{\cdot}6 \mp 3(0{\cdot}0245)$$

i.e., 0·5265 and 0·6735

i.e., 0·53 and 0·67.

Ex. 15-9. *A dealer takes* 100 *samples from a consignment of* 1000 *items of a certain good and finds that there are* 50 *items of grade I worth Rs.* 5 *per thousand,* 30 *items of grade II worth Rs.* 4 *per thousand and* 20 *items of grade III worth Rs.* 3 *per thousand. Within what limits should the value of the consignment be fixed ?*

Sol. *Grade I.*

$$\text{Proportion of items} = \frac{50}{100} = 0{\cdot}5$$

$\therefore$ S.E. of the proportion of items

$$= \sqrt{\frac{1}{100}(0{\cdot}5)\,(0{\cdot}5)} = 0{\cdot}05$$

$\therefore$ Probable limits to the proport

$0{\cdot}5 \mp 3(0{\cdot}05)$ *i.e.*, 0·35 ai

Grade II

$$\text{Proportion of items} = \frac{30}{100} = 0{\cdot}3$$

$\therefore$ S.E. of the proportion

$= \sqrt{\ }$

$\therefore$ Probable limits to the propor

$0{\cdot}3 \mp 3(0{\cdot}0458)$ *i.e.*, 0·16

Grade III

$$\text{Proportion of items} = \frac{20}{100} = 0{\cdot}2$$

$\therefore$ S.E. of the proportion

$= \sqrt{\ }$

$\therefore$ Probable limits to the propoı

$0{\cdot}2 \mp 3(0{\cdot}04)$ *i.e.*, 0·08 a

Now the highest value that can grade I is the highest and grade III tł

Proportion of grade I = 0·65

and proportion of grade III = 0·08

$\therefore$ Proportion of grade II = 1

$\therefore$ Highest value of the consi

= (0

The lowest value that can be giv the lowest the grade III the highest s

Proportion of grade I = 0·35

and proportion of grade III = 0·32

$\therefore$ Proportion of grade II = 1

$\therefore$ The least value of the con

= (0

$\therefore$ The limits of the value of

1. A coin is tossed 10,000 times a that the coin is unbiased ?
2. In 324 throws of a six-faced die as unbiased ?
3. In breeding certain stocks, 40 expectation is one-fourth glal occurred as a fluctuation of sa
4. Experience has shown that 10 day's production of 400 articl hypothesis of 10 percent?
5. Balls are drawn from a bag c replacement. In 2000 drawing bias in the drawer ?

$\therefore$ Probable limits to the proportion of items are

$$0 \cdot 5 \mp 3(0 \cdot 05) \text{ i.e., } 0 \cdot 35 \text{ and } 0 \cdot 65$$

Grade II

$$\text{Proportion of items} = \frac{30}{100} = 0 \cdot 3$$

$\therefore$ S.E. of the proportion

$$= \sqrt{\frac{(0 \cdot 3)(0 \cdot 7)}{100}} = 0 \cdot 0458$$

$\therefore$ Probable limits to the proportion are

$$0 \cdot 3 \mp 3(0 \cdot 0458) \text{ i.e., } 0 \cdot 1626 \text{ and } 0 \cdot 4374.$$

Grade III

$$\text{Proportion of items} = \frac{20}{100} = 0 \cdot 2$$

$\therefore$ S.E. of the proportion

$$= \sqrt{\frac{1}{100}(0 \cdot 2)(0 \cdot 8)} = 0 \cdot 04$$

$\therefore$ Probable limits to the proportion are

$$0 \cdot 2 \mp 3(0 \cdot 04) \text{ i.e., } 0 \cdot 08 \text{ and } 0 \cdot 32.$$

Now the highest value that can be given to the consignment is that value for which grade I is the highest and grade III the lowest so that

Proportion of grade I $= 0 \cdot 65$

and proportion of grade III $= 0 \cdot 08$

$\therefore$ Proportion of grade II $= 1 - 0 \cdot 65 - 0 \cdot 08 = 0 \cdot 27$

$\therefore$ Highest value of the consignment

$$= (0 \cdot 65)(5) + (0 \cdot 27)(4) + (0 \cdot 08)(3) = \text{Rs. } 4 \cdot 57.$$

The lowest value that can be given to the consignment is that value for which grade I is the lowest the grade III the highest so that

Proportion of grade I $= 0 \cdot 35$

and proportion of grade III $= 0 \cdot 32$

$\therefore$ Proportion of grade II $= 1 - 0 \cdot 35 - 0 \cdot 32 = 0 \cdot 33$

$\therefore$ The least value of the consignment

$$= (0 \cdot 35)(5) + (0 \cdot 33)(4) + (0 \cdot 32)3 = \text{Rs. } 4 \cdot 03.$$

$\therefore$ The limits of the value of the consignment are Rs. $4 \cdot 03$ and Rs. $4 \cdot 57$.

EXERCISES

1. A coin is tossed 10,000 times and it turns up head 5195 times. Is it reasonable to think that the coin is unbiased ? **[Ans.** No]
2. In 324 throws of a six-faced die odd points appeared 181 times. Can the die be regarded as unbiased ? **[Ans.** Insignificant at 1% level]
3. In breeding certain stocks, 408 hairy and 126 glabrous plant were obtained. If the expectation is one-fourth glabrous, is the divergence significant or might it have occurred as a fluctuation of sampling ? **[Ans.** Insignificant]
4. Experience has shown that 10% of a manufactured product is of top quality. In one day's production of 400 articles only 50 are of top quality. Does this contradict our hypothesis of 10 percent? **[Ans.** No]
5. Balls are drawn from a bag containing equal number of black and white balls with replacement. In 2000 drawings, 1100 black and 900 white balls appear. Is there some bias in the drawer ? **[Ans.** Yes]

6. A personnel manager claims that 80% of all single woman hired for secretarial job get married and quit work within two years after they are hired. Test this hypothesis at 5% level of significance, if among 200 such secretaries 112 got married within two years after they were hired and quit their jobs. **[Ans.** Hypothesis is wrong]
7. 400 apples are taken from a large consignment and 50 are found to be bad. Estimate the percentage of bad apples in the consignment and assign the limits within which the percentage lies. **[Ans.** 7·5% and 17·5%]
8. Given that, on the average 40 out of 1000 insured men of age 60 die within a year and that 60 of a particular group of 1000 such men died within a year, show that this group cannot be regarded as representative sample, seeing that the actual deviation of the proportion of deaths is more than three times the standard error of the proportion for samples of the size.
9. A man buys 100 sacks of tonatoes. He finds that out of 100 tomatoes chosen from the sacks at random, 40 are of type A, worth Rs. 10 a sack, 25 are of type B, worth Rs. 7 a sack, 20 are of type C, worth Rs. 5 a sack and 15 are of class D, worth Rs. 4 per sack. What are the upper and lower limits for the value of the tomatoes ? **[Ans.** 634·78 and 835·21]
10. 12 dice were thrown 6500 times; 4, 5 or 6 being reckoned as a success. What proportion of success do you expect ? If in actual observation the proportion of success is found to be 0·5016, find the standard deviation of proportion with the given number of throws and state whether you would regard the excess of successes as probably significant bias in the dice.
11. In a certain maternity home during a year there were 1600 births of which 840 were males. Test the hypothesis that male and female births are equally likely. Supposing the null hypothesis is not given, determine the $\pm 3\sigma$ confidence limits for the proportion of male births.
12. A biased coin was thrown 400 times and heads resulted 240 times. Show that the probability of throwing head in a single trial almost certainly lies between 0·53 and 0·67.
13. In a sample of 500 people in Kerala 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in this state at 5% level of significance ? **[Ans.** Yes]
14. A random sample of 16 values from a normal population showed a mean of 41·5 inches and a sum of squares of deviation from this mean equal to 135 (inch)^2. Show that the assumption of a mean of 43·5 inches for the population is not reasonable and that the 95% confidence limits for this mean are 39·9 and 43·1 inches.
15. Show that the prob. that the number of heads in 400 throws of a fair coin lies between 180 and 220 is approximately $2F(2) - 1$ where $F(x)$ denotes the standard normal distribution function *i.e.*,

$$F(x) = P(X \leq x), \; X \sim N(0, 1)$$

15.2-2. Comparison of Large Samples

Let the two populations be tested for the prevalence of a certain attribute A by taking from them large simple samples of sizes n_1 and n_2 respectively. Let x_1 and x_2 be the number of individuals possessing A in the two samples.

Let $$p_1 = \frac{x_1}{n_1} \quad \text{and} \quad p_2 = \frac{x_2}{n_2}$$

Let P_1 and P_2 be probabilities for an individual to possess A for two populations.

Then $$E(x_1) = n_1 P_1 \text{ and hence } E(p_1) = P_1$$

$$\text{Var}(x_1) = n_1 P_1 Q_1 \text{ and hence var}(p_1) = \frac{P_1 Q_1}{n_1}, \; Q_1 = 1 - P_1$$

Similarly, $E(p_2) =$

Now $E(p_1 - p_2) =$

and var $(p_1 - p_2) =$

$\therefore$ $u =$

is approximately a $N(0, 1)$.

(*i*) The hypothesis to be teste

'Is the difference $(p_1 - p_2)$ sig

w.r.t. A'.

To proceed with we set the hy

basis of this hypothesis

$P_1 =$

$\therefore$ $u =$

where $Q = 1 - P$

we now test the significance with t

Thus if (*i*) $|u| < 1{\cdot}96$, the hypo

(*ii*) $1{\cdot}96 < |u| < 2{\cdot}58$, the diffe

(*iii*) $2{\cdot}58 < |u| < 3$, the differe

(*iv*) $|u| > 3$, the hypothesis is no

Generally P is unknown so w

unbiased estimate of P is given by

$P =$

It is unbiased because

$E(P) =$

$=$

(*ii*) The hypothesis to be teste

'Is the real difference between t

i.e., if in populations $P_2 < P_1$, is it l

i.e., $p_1 - p_2 \leq$ (

i.e., $(P_1 - P_2) + eu \leq$ (

i.e., $u \leq$ -

where $e =$

Now $P(u > 1{\cdot}645) =$ C

woman hired for secretarial job get
re hired. Test this hypothesis at 5%
s 112 got married within two years
[**Ans.** Hypothesis is wrong]
d 50 are found to be bad. Estimate
and assign the limits within which
[**Ans.** 7·5% and 17·5%]
men of age 60 die within a year and
died within a year, show that this
e, seeing that the actual deviation of
the standard error of the proportion

out of 100 tomatoes chosen from the
sack, 25 are of type B, worth Rs. 7
are of class D, worth Rs. 4 per sack.
e of the tomatoes ?
[**Ans.** 634·78 and 835·21]
koned as a success. What proportion
n the proportion of success is found
oportion with the given number of
e excess of successes as probably

were 1600 births of which 840 were
births are equally likely. Supposing
σ confidence limits for the proportion

s resulted 240 times. Show that the
st certainly lies between 0·53 and 0·67.
inkers and the rest are coffee drinkers.
lly popular in this state at 5% level of
[**Ans.** Yes]
l population showed a mean of 41·5
this mean equal to 135 (inch)^2. Show
r the population is not reasonable and
re 39·9 and 43·1 inches.
400 throws of a fair coin lies between
re $F(x)$ denotes the standard normal

0, 1)

ence of a certain attribute A by taking
spectively. Let x_1 and x_2 be the number

$\dfrac{x_2}{n_2}$
to possess A for two populations.
$E(p_1) = P_1$

e var $(p_1) = \dfrac{P_1Q_1}{n_1}$, $Q_1 = 1 - P_1$

Similarly, $$E(p_2) = P_2 \text{ and } \operatorname{var}(p_2) = \frac{P_2Q_2}{n_2},\ Q_2 = 1 - P_2$$

Now $$E(p_1 - p_2) = E(p_1) - E(p_2) = P_1 - P_2$$

and $$\operatorname{var}(p_1 - p_2) = \operatorname{var}(p_1) + \operatorname{var}(p_2) = \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}$$

$$\therefore \quad u = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}}$$

is approximately a $N(0, 1)$.

(*i*) The hypothesis to be tested is :

'Is the difference $(p_1 - p_2)$ significant of a real difference between the two populations *w.r.t. A*'.

To proceed with we set the hypothesis that two populations are similar *w.r.t. A*. On the basis of this hypothesis

$$P_1 = P_2 = P \text{ (say)}$$

$$\therefore \quad u = \frac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

where $Q = 1 - P$

we now test the significance with the aid of the normal curve.

Thus if (*i*) $|u| < 1·96$, the hypothesis is acceptable at 5% level of significance.

(*ii*) $1·96 < |u| < 2·58$, the difference is significant at 5% level of significance.

(*iii*) $2·58 < |u| < 3$, the difference is significant at 1% level of significance.

(*iv*) $|u| > 3$, the hypothesis is not acceptable and hence the difference is highly significant.

Generally P is unknown so we have to estimate it from the sample proportions. An unbiased estimate of P is given by

$$P = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

It is unbiased because

$$E(P) = \frac{E(x_1 + x_2)}{n_1 + n_2} = \frac{n_1P_1 + n_2P_2}{n_1 + n_2}$$

$$= P \qquad (\because\ P_1 = P_2 = P)$$

(*ii*) The hypothesis to be tested is :

'Is the real difference between the populations likely to be hidden in two samples drawn' *i.e.*, if in populations $P_2 < P_1$, is it likely that p_1 and p_2 will be s.t. $p_1 \le p_2$

i.e., $$p_1 - p_2 \le 0$$

i.e., $$(P_1 - P_2) + eu \le 0$$

i.e., $$u \le -\frac{(P_1 - P_2)}{e}$$

where $$e = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

Now $$P(u > 1·645) = 0·5 - \int_0^{1·645} dP = 0·5 - 0·45 = 0·05$$

$$\therefore \quad P(u < -1.645) = 0.05$$

$$\therefore \text{ If } \quad \frac{P_1 - P_2}{e} > 1.645$$

$$i.e., \quad -\frac{(P_1 - P_2)}{e} < -1.645$$

$$P\left(u \le -\frac{(P_1 - P_2)}{e}\right) < 0.05$$

which implies that at 5% level, it is unlikely that difference will be hidden in simple sampling.

Similarly, if $\dfrac{P_1 - P_2}{e} > 2.327$, the difference is unlikely to be hidden at 1% level of significance.

Ex. 15-10. *In two large populations there are* 30% *and* 25% *respectively of fair haired people. Is this difference likely to be hidden in samples of* 1200 *and* 900 *respectively from the two populations ?*

Sol. Here, $P_1 = 0.3$, $P_2 = 0.25$, $n_1 = 1200$ and $n_2 = 900$

$$\therefore \quad e = \sqrt{\frac{(0.3)(0.7)}{1200} + \frac{(0.25)(0.75)}{900}} \cong 0.0195$$

$$\therefore \quad \frac{P_1 - P_2}{e} = \frac{0.05}{0.0195} \cong 2.56 \ (> 1.645)$$

$\therefore$ At 5% level the difference is unlikely to be hidden.

Ex. 15-11. *In a simple sample of* 600 *men from a certain large city,* 400 *are found to be smokers. In one of* 900 *from another large city,* 450 *are smokers. Do the data indicate that the cities are significantly different with respect to the prevalence of smoking among men ?*

Sol. Set the hypothesis : Two cities do not differ significantly *w.r.t.* the prevalence of smoking among men.

$$n_1 = 600, \qquad x_1 = 400$$

$$n_2 = 900, \qquad x_2 = 450$$

$$\therefore \quad p_1 = \frac{2}{3} \quad \text{and} \quad p_2 = \frac{1}{2}$$

$$\therefore \quad P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{850}{1500} = \frac{17}{30}$$

$$\therefore \quad Q = 1 - P = \frac{13}{30}$$

$$\therefore \quad e^2 = PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{17}{30}\,\frac{13}{30}\left(\frac{1}{600} + \frac{1}{900}\right)$$

$$= 0.000682$$

$$\therefore \quad e = 0.026$$

$$\therefore \quad u = \frac{p_1 \sim p_2}{e} = \frac{\frac{2}{3} - \frac{1}{2}}{0.026} = 6.4 \ (> 3)$$

$\therefore$ The difference is highly significant and hence the two cities are significantly different with respect to the prevalence of smoking habit among men.

Ex. 15-12. *A railway company installed two sets of* 50 *Burmaties each. The two sets were treated with creosate by two different processes. After a number of years of service it*

was found that 22 ties of first set and 18 ties of the second set were still in good condition. Are we justified in claiming that there is no real difference between the preserving properties of the two processes ?

Sol. Set the hypothesis : There is no real difference between the preserving properties of two processes.

Here
$$p_1 = 0{\cdot}44,\ p_2 = 0{\cdot}36$$
$$P = 0{\cdot}4,\ Q = 0{\cdot}6 \text{ and } e = 0{\cdot}098$$
$$\therefore \quad u \simeq 0{\cdot}8 < 1{\cdot}96$$

$\therefore$ Data provides no evidence against the hypothesis.

Ex. 15-13. *In a referendum submitted to the student body at a university 850 men and 566 women voted. 530 of the men and 304 of the women voted yes. Does this indicate a significant difference of opinion on the matter, at the 1% level, between men and women students ?*

Sol. Set the hypothesis : There is no significant difference of opinion between men and women on the matter.

Here
$$p_1 = 0{\cdot}6235,\quad p_2 = 0{\cdot}5371$$
$$P = \frac{834}{1416},\quad Q = \frac{582}{1416} \quad \text{and} \quad e = 0{\cdot}0267$$
$$\therefore \quad u = 3{\cdot}2\ (> 3)$$

$\therefore$ Hypothesis is wrong.

Ex. 15-14. *On the basis of their total scores, the 200 candidates at a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of this examination. Among the first group, 40 had the correct answer; whereas among the second group 80 had the correct answer. On the basis of these results, can one conclude that the first question is no good at discriminating ability of the type being examined here ?*

Sol. Set the hypothesis : The first question is no good at discriminating ability of the type being examined.

$$\text{Here } n_1 = \frac{30}{100} \cdot 200 = 60,\ n_2 = 140,\ x_1 = 40, \text{ and } x_2 = 80$$
$$\therefore \quad p_1 = 0{\cdot}6667 \text{ and } p_2 = 0{\cdot}5714$$
$$P = 0{\cdot}6,\quad Q = 0{\cdot}4 \quad \text{and} \quad e = \frac{1}{5\sqrt{7}}$$
$$\therefore \quad u \simeq 1{\cdot}26\ (< 1{\cdot}96)$$

$\therefore$ The data is consistent with the hypothesis.

Ex. 15-15. *In a year there are 956 births in a town A of which 52·5% were males, while in towns A and B combined this proportion in a total of 1406 births was 0·496. Is there any significant difference in the proportion of male births in the two towns ?*

Sol. Set the hypothesis : There is no significant difference in the proportion of male births in the two towns.

Here
$$n_1 = 956,\qquad n_1 + n_2 = 1406 \qquad \therefore \quad n_2 = 450$$
$$p_1 = 0{\cdot}525,\qquad P = 0{\cdot}496 \qquad \therefore \quad Q = 0{\cdot}504$$
$$\therefore \quad p_2 = 0{\cdot}434$$
$$\therefore \quad u \simeq 3{\cdot}2\ (> 3)$$

$\therefore$ Hypothesis is wrong.

Ex. 15-16. *A machine puts out 16 imperfect articles in a sample of 500. After machine is overhauled it puts out 3 imperfect articles in a batch of 100. Has the machine been improved ?*

Sol. Set the hypothesis : Machine has not been improved.

Here $\qquad p_1 = 0{\cdot}032,\ p_2 = 0{\cdot}03,\ P = \dfrac{19}{600},\ Q = \dfrac{581}{600}$

$\therefore \qquad u \cong 0{\cdot}1\ (< 1{\cdot}96)$

$\therefore$ Hypothesis may be correct.

Ex. 15-17. *In a large city A, 20% of a random sample of 900 school boys had a certain slight physical defect. In another large city B 18·5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant ?*

Sol. Here $\qquad p_1 = 0{\cdot}2,\quad p_2 = 0{\cdot}185,\quad P = 0{\cdot}19,\ Q = 0{\cdot}81$

$\therefore \qquad u \cong 0{\cdot}92\ (< 1{\cdot}96)$

$\therefore$ Difference is not significant.

Ex. 15-18. (*a*) *Two large random samples of sizes n_1 and n_2 are taken from two populations. If p_1 and p_2 be the proportions of members possessing the attribute in two samples, give procedure of testing the significance of the difference between p_1 and*

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

(*b*) *In two random samples of 400 and 500 students from two different colleges,* 300 *students in each were found to be failed in an examination. Find out whether the proportion of failures in first college is significantly greater than the proportion of failures in two colleges taken together.*

Sol. (*a*) Let $\bar{p}_1$ and $\bar{p}$ be the expected values of p_1 and p. Then

$$\bar{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2}$$

where $\bar{p}_2 = E(p_2)$

Now
$$\begin{aligned}
\text{cov}\,(p_1, p) &= E(p_1 - \bar{p}_1)(p - \bar{p}) \\
&= \frac{1}{n_1 + n_2} E[(p_1 - \bar{p}_1)\{n_1(p_1 - \bar{p}_1) + n_2(p_2 - \bar{p}_2)\}] \\
&= \frac{1}{n_1 + n_2} \{n_1 E(p_1 - \bar{p}_1)^2 + n_2 E(p_1 - \bar{p}_1)(p_2 - \bar{p}_2)\} \\
&= \frac{n_1}{n_1 + n_2} \text{ var}\,(p_1)
\end{aligned}$$

$(\because\ \text{cov}\,(p_1, p_2) = 0$ as p_1, p_2 are independent)

Now p gives the estimate of population proportion and hence var $(p_1) = \dfrac{pq}{n_1}$ and var $(p_2) = \dfrac{pq}{n_2}$

$$\therefore \qquad \text{cov}\,(p_1, p) = \frac{pq}{n_1 + n_2}$$

Now
$$\begin{aligned}
\text{var}\,(p) &= \frac{1}{(n_1 + n_2)^2} \text{ var}\,(n_1 p_1 + n_2 p_2) \\
&= \frac{1}{(n_1 + n_2)^2} \{n_1^2 \text{ var}\,(p_1) + n_2^2 \text{ var}\,(p_2)\} \\
&= \frac{pq}{n_1 + n_2}
\end{aligned}$$

$$\therefore \qquad \text{var}\,(p - p_1) = \text{var}\,(p) + \text{var}\,(p_1) - 2\,\text{cov}\,(p_1, p)$$

$\frac{19}{600}$, $Q=\frac{581}{600}$

of 900 school boys had a certain
random sample of 1600 school
oportions significant ?
$0{\cdot}19,\ Q=0{\cdot}81$

s n_1 and n_2 are taken from two
s possessing the attribute in two
difference between p_1 and

from two different colleges, 300
. Find out whether the proportion
the proportion of failures in two

and p. Then

$_1(p_1-\bar{p}_1)+n_2(p_2-\bar{p}_2)\}]$

$)^2+n_2E(p_1-\bar{p}_1)(p_2-\bar{p}_2)\}$

$p_2)=0$ as p_1, p_2 are independent)
and hence var $(p_1)=\frac{pq}{n_1}$ and var

$_1+n_2p_2)$

$(p_1)+n_2^2$ var $(p_2)\}$

2 cov (p_1, p)

$$= \frac{pq}{n_1+n_2} + \frac{pq}{n_1} - \frac{2pq}{n_1+n_2}$$

$$= \frac{n_2}{n_1}\cdot\frac{pq}{n_1+n_2}$$

Assuming the hypothesis that there is no significant difference between p_1 and p, the test statistic becomes

$$u = \frac{p_1-p}{\sqrt{\dfrac{n_2}{n_1}\cdot\dfrac{pq}{n_1+n_2}}}$$

which is a $N(0, 1)$ as n_1 and n_2 are large.

(*b*) Here $n_1=400, n_2=500, p_1=\frac{3}{4}$ and $p_2=\frac{3}{5}$

$$\therefore \quad p = \frac{300+300}{400+500} = \frac{2}{3}$$

$$\therefore \quad q = 1-\frac{2}{3} = \frac{1}{3}$$

$$\therefore \quad u = \frac{\frac{3}{4}-\frac{2}{3}}{\sqrt{\dfrac{5}{4}\cdot\dfrac{\frac{2}{3}\cdot\frac{1}{3}}{900}}} = 4{\cdot}7\ (>3)$$

$\therefore$ p and p_1 are significantly different. Obviously $p_1>p$.

EXERCISES

1. In a random sample of 500 persons from town A, 200 are found to be smokers. In a sample of 400 from town B, 200 are found to be smokers. Discuss the question whether the data reveal a significant difference between A and B so far as the smoking habit among persons is concerned. **[Ans.** Significant]
2. In a large city A, 20% of a random sample of 900 schoolboys had defective eye-sight. In another large city B, 15·5% of a random sample of 1600 schoolboys had the same defect. Is the difference between the two proportions significant? **[Ans.** Not Significant]
3. In a random sample of 500 men from a particular district, 300 are found to be smokers. In one of 1000 men from another district, 550 are smokers. Do the data indicate that the two districts are significantly different w.r.t the prevalence of smoking among men? **[Ans.** Not Significant]
4. From each of two consignments of eggs, a sample of size 200 is drawn and the number of rotton eggs counted. Test whether the proportion of rotton eggs in the two consignments are significantly different or not, given that

	Size of sample	*No. of rotton eggs*
Sample from consignment A	200	40
Sample from consignment B	200	30

[Ans. Not Significant]

5. In two large populations there are 35% and 30% of fair haired people. Is the difference likely to be hidden in simple samples of 1500 and 1000 ? **[Ans.** $e=0{\cdot}019$, unlikely]

6. A machine produces 30 defective screws in a lot of 1000. After overhauling it produced 20 defective in a lot of 800. Set-up a statistical hypothesis and test it. [**Ans.** $u = 0{\cdot}6$]

Ex. 15-19. *Show that standard error of the number of success is the square root of the mean number of successes provided the mean proportion of successes is small.*

Sol. Mean proportion of successes $= p$

$\therefore$ It p is small, standard error of the number of successes

$$= \sqrt{npq} = \sqrt{np(1-p)}$$
$$\simeq \sqrt{np}$$
$$= \sqrt{\text{Mean number of successes}}.$$

Ex. 15-20. *Show that precision of the proportion of successes varies as the square root of the number of members in the sample.*

Ex. 15-21. *If for one half of n events, the chance of success is p and the chance of failure is q, whilst for the other half the chance of success is q and the chance of failure is p. Show that the standard deviation of the number of successes is the same as if the chance of success were p in all the cases i.e., $\sqrt{npq}$ but that the mean of the number of successes is $\frac{n}{2}$ and not np.*

Sol. Let x_1 and x_2 denote the number of successes in two halves.

Then $$E(x_1) = \frac{n}{2}\,p,\ \text{var}(x_1) = \frac{n}{2}\,pq$$

and $$E(x_2) = \frac{n}{2}\,q,\ \text{var}(x_2) = \frac{n}{2}\,pq$$

$$\therefore \quad E(x_1 + x_2) = \frac{n}{2}(p+q) = \frac{n}{2}$$

$$\text{var}(x_1 + x_2) = \text{var}(x_1) + \text{var}(x_2) \qquad [\because \text{ The halves are independent}]$$

$$= \frac{n}{2}\,pq + \frac{n}{2}\,pq = npq.$$

Ex. 15-22. *The sex ratio at birth is sometimes given by the ratio of male to female births, instead of the proportion of male to total births. If z is the ratio i.e., $z = \frac{p}{q}$, show that the standard error of z is approximately $\frac{1}{1+z}\sqrt{\frac{z}{n}}$, n being large so that deviations are small compared with the mean.*

Sol. Let x be the number of male births. Then $(n-x)$ is the number of female births.

Now $$z = \frac{p}{q} = \frac{p}{1-p}$$

$$\therefore \quad p = \frac{z}{1+z} \text{ and } q = \frac{1}{1+z}$$

Also $$z = \frac{x}{n-x} = \frac{x}{n}\cdot\left\{1-\frac{x}{n}\right\}^{-1} = \frac{x}{n}\left(1+\frac{x}{n}+\ldots\ldots\right) \simeq \frac{x}{n}$$

$$\therefore \quad \text{S.E. of } z \simeq \text{S.E. of}\left(\frac{x}{n}\right) \simeq \sqrt{\frac{pq}{n}}$$

$$\simeq \frac{1}{1+z}\sqrt{\frac{z}{n}}.$$

Ex. 15-23. *n individuals fall into one or the other two categories with probabilities p and $q(=1-p)$, the number in two categories being n_1 and n_2. Show that $\text{cov}(n_1, n_2) = -npq$.*

Hence obtain var $\left(\frac{n_1}{n} - \frac{n_2}{n}\right)$.

Sol. Evidently n_2

Now $E(n_1)$

$\therefore$ $E(n_2)$

$\therefore$ $\text{cov}(n_1, n_2)$

$\text{var}\left(\frac{n_1}{n} - \frac{n_2}{n}\right)$

15.3. Sample Mean

A sample of size n can be

....., x_n denote the random variabl

variable defined by

$\bar{x}$

The probability distribution

or the sampling distribution of

Remark. In the case of rar

15.3.1. Central Limit Theorer

Statement. If $x_1, x_2, \ldots\ldots, x_n$ b

mean μ and s.d. σ and $\bar{x} = \frac{1}{n}\sum$

has distribution that approaches

Proof. $M_0(t)$ of z =

=

=

=

000. After overhauling it produced
pothesis and test it. [**Ans.** $u = 0{\cdot}6$]
of success is the square root of the
of successes is small.

cesses

ccesses.
successes varies as the square root

of success is p and the chance of
is q and the chance of failure is p.
esses is the same as if the chance of
ean of the number of successes is $\frac{n}{2}$

in two halves.

q

q

[$\because$ The halves are independent]

q.

given by the ratio of male to female
If z is the ratio i.e., $z = \frac{p}{q}$, show that
n being large so that deviations are

$n - x$) is the number of female births.

$$\Big\}^{-1} = \frac{x}{n}\left(1+\frac{x}{n}+\ldots\right) \simeq \frac{x}{n}$$

$$\frac{pq}{n}$$

her two categories with probabilities p
n_1 and n_2. Show that $\text{cov}(n_1, n_2) = -npq$.

Hence obtain var $\left(\frac{n_1}{n} - \frac{n_2}{n}\right)$.

Sol. Evidently $\quad n_2 = n - n_1$

Now $\quad E(n_1) = np$

$$\therefore \quad E(n_2) = n - E(n_1) = n - np = nq$$

$$\begin{aligned}
\therefore \quad \text{cov}(n_1, n_2) &= E\{(n_1 - np)(n_2 - nq)\} \\
&= E\{(n_1 - np)(n - n_1 - nq)\} \\
&= -E(n_1 - np)^2 = -\text{var}(n_1) = -npq
\end{aligned}$$

$$\begin{aligned}
\text{var}\left(\frac{n_1}{n} - \frac{n_2}{n}\right) &= \frac{1}{n^2}\cdot \text{var}(n_1 - n_2) \\
&= \frac{1}{n^2}[\text{var}(n_1) + \text{var}(n_2) - 2\,\text{cov}(n_1, n_2)] \\
&= \frac{1}{n^2}[npq + npq + 2npq] \\
&= \frac{4pq}{n}.
\end{aligned}$$

15.3. Sample Mean

A sample of size n can be described by the values of the random variables. Let $x_1, x_2, \ldots, x_n$ denote the random variables for a sample of size n. Then the sample mean is a random variable defined by

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

The probability distribution of $\bar{x}$ is called sampling distribution for the sample mean $\bar{x}$ or the sampling distribution of mean.

Remark. In the case of random sample, $x_1, x_2, \ldots, x_n$ are independent.

15.3.1. Central Limit Theorem

Statement. If $x_1, x_2, \ldots, x_n$ be n independent random variables all with same distribution, mean μ and s.d. σ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, then, if *m.g.f.* of x_1 exist, the variate $z = \frac{(\bar{x} - \mu)}{\sigma}\sqrt{n}$ has distribution that approaches the standard normal distribution as $n \to \infty$.

Proof.

$$\begin{aligned}
M_0(t) \text{ of } z &= E\left\{e^{t\frac{(\bar{x}-\mu)}{\sigma}\sqrt{n}}\right\} \\
&= e^{-t\frac{\mu\sqrt{n}}{\sigma}}\, E\left\{e^{t\frac{\sqrt{n}}{\sigma}\bar{x}}\right\} \\
&= e^{-t\frac{\mu\sqrt{n}}{\sigma}}\, E\left\{e^{\frac{t}{\sigma\sqrt{n}}(x_1 + x_2 + \ldots + x_n)}\right\} \\
&= e^{-t\frac{\mu\sqrt{n}}{\sigma}}\, E\left\{e^{\frac{t}{\sigma\sqrt{n}}x_1}\, e^{\frac{t}{\sigma\sqrt{n}}x_2} \ldots\, e^{\frac{t}{\sigma\sqrt{n}}x_n}\right\}
\end{aligned}$$

$$= e^{-t\frac{\mu\sqrt{n}}{\sigma}} E\left\{e^{\frac{t}{\sigma\sqrt{n}}x_1}\right\} E\left\{e^{\frac{t}{\sigma\sqrt{n}}x_2}\right\} \ldots\ldots E\left\{e^{\frac{t}{\sigma\sqrt{n}}x_n}\right\}$$

(as x's are independent)

Now since all x_i has same distribution, mean and s.d. their m.g.f. will also be same.

$$\therefore \quad M_0(t) \text{ of } z = e^{-t\frac{\mu\sqrt{n}}{\sigma}} \left\{M_0\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n$$

where $M_0\left(\frac{t}{\sigma\sqrt{n}}\right)$ is the m.g.f. of x_i.

$$\therefore \quad \log\{M_0(t) \text{ of } z\} = -t\frac{\mu\sqrt{n}}{\sigma} + n\log\left\{M_0\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}$$

$$= -t\frac{\mu\sqrt{n}}{\sigma} + n\log\left\{1+\mu_1'(0)\frac{t}{\sigma\sqrt{n}} + \frac{\mu_2'(0)}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \ldots\right\}$$

$$= -t\frac{\mu\sqrt{n}}{\sigma} + n\left\{\left\{\mu_1'(0)\frac{t}{\sigma\sqrt{n}} + \frac{\mu_2'(0)}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \ldots\right\} - \frac{1}{2}\left\{\mu_1'(0)\frac{t}{\sigma\sqrt{n}} + \ldots\right\}^2 + \ldots\right\}$$

Now $\mu'_1(0) = \mu$.

$\therefore \quad \log\{M_0(t) \text{ of } z\} = \frac{t^2}{2\sigma^2}[\mu_2'(0) - \{\mu_1'(0)\}^2]$ + terms containing n in the denominator.

$$\therefore \quad \lim_{n\to\infty} \log\{M_0(t) \text{ of } z\} = \frac{t^2}{2\sigma^2}[\mu'_2(0) - \{\mu'_1(0)\}^2] = \frac{t^2}{2}$$

$$[\because \mu'_2(0) - \{\mu'_1(0)\}^2 = \mu_2 = \sigma^2]$$

$$\therefore \quad \lim_{n\to\infty} M_0(t) \text{ of } z = e^{\frac{1}{2}t^2}$$

which is the m.g.f. of a $N(0, 1)$

$\therefore$ As $n \to \infty$ the distribution of z tends to the standard normal distribution.

15.3-2. The standard error of the mean of a random sample of size n from a population with variance σ^2

Sol. Let $x_1, x_2, \ldots, x_n$ be a random sample.

Then
$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

is the sample mean.

$$\therefore \quad \text{var}(\bar{x}) = \frac{1}{n^2}[\text{var}(x_1 + x_2 + \ldots + x_n)]$$

$$= \frac{1}{n^2}[\text{var}(x_1) + \text{var}(x_2) + \ldots + \text{var}(x_n)]$$

(as x's are independent)

$= \frac{\sigma}{}$

$\therefore$ S.E. of $\bar{x}$ $= \frac{}{\sqrt{}}$

15.4. Sampling of Variables

In this case population is the freq provides a value of the variable. Dra the variable from those of the distrib

15.4-1. Unbiased Estimate of Popu

Let the values $X_1, X_2, \ldots, X_N$ cons

Then $\mu = \frac{X}{}$

Let $x_1, x_2, \ldots, x_n$ be a random sa

$\bar{x} = \frac{1}{n}$

$\therefore$ $E(\bar{x}) = \frac{1}{n}$

Now for fixed i, x_i can take any

$\frac{1}{N}$.

$\therefore$ $E(x_i) = \frac{1}{N}$

$\therefore$ $E(\bar{x}) = \frac{1}{N}$

$\therefore$ Sample mean $\bar{x}$ is an unbi

15.4-2. Unbiased Estimate of Popul

Sample s.d. is

$s^2 = \frac{1}{n}$

$= \frac{1}{n}$

$= \frac{1}{n}$

$= \frac{1}{n}$

$\therefore$ $E(s^2) = \frac{1}{n}$

Since $E(x_i) = \mu =$

and $E(\bar{x} - \mu)^2 =$ var

$\therefore$ $E(s^2) = \sigma^2 -$

$$\left\{e^{\frac{t}{\sigma\sqrt{n}}x_2}\right\} \ldots\ldots E\left\{e^{\frac{t}{\sigma\sqrt{n}}x_n}\right\}$$

(as x's are independent)

their m.g.f. will also be same.

$$\frac{t}{\sigma\sqrt{n}} + \frac{\mu_2'(0)}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \ldots\ldots\Big\}$$

$$+ \frac{\mu_2'(0)}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \ldots\ldots\Big\}$$

$$\ldots\ldots\Big\}$$

] + terms containing n in the

$$0)\}^2] = \frac{t^2}{2}$$

$\because\ \mu_2'(0) - \{\mu_1'(0)\}^2 = \mu_2 = \sigma^2]$

dard normal distribution.

ample of size n from a population

$. + x_n)]$

$_2) + \ldots + \mathrm{var}(x_n)]$

(as x's are independent)

$$= \frac{\sigma^2}{n} \qquad (\because \mathrm{var}(x_i) = \sigma^2 \text{ for all } `i')$$

$$\therefore \qquad \text{S.E. of } \bar{x} = \frac{\sigma}{\sqrt{n}}.$$

15.4. Sampling of Variables

In this case population is the frequency distribution of the variable and its each member provides a value of the variable. Drawing of a sample is same as choosing certain values of the variable from those of the distribution.

15.4-1. Unbiased Estimate of Population Mean

Let the values $X_1, X_2, \ldots, X_N$ constitute a finite population with mean μ and variance σ^2.

Then
$$\mu = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

Let $x_1, x_2, \ldots, x_n$ be a random sample from the population. The sample mean is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\therefore \qquad E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$$

Now for fixed i, x_i can take any one of the values $X_1, X_2, \ldots, X_N$ each with probability $\frac{1}{N}$.

$$\therefore \qquad E(x_i) = \frac{1}{N}(X_1 + X_2 + \ldots X_N) = \mu$$

$$\therefore \qquad E(\bar{x}) = \frac{1}{N}\sum_{i=1}^{n} \mu = \mu$$

$\therefore$ Sample mean $\bar{x}$ is an unbiased estimate of the population mean.

15.4-2. Unbiased Estimate of Population Variance

Sample s.d. is

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu + \mu - \bar{x})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\{(x_i - \mu)^2 + (\mu - \bar{x})^2 + 2(\mu - \bar{x})(x_i - \mu)\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 - (\mu - \bar{x})^2$$

$$\therefore \qquad E(s^2) = \frac{1}{n}\sum_{i=1}^{n}E(x_i - \mu)^2 - E(\bar{x} - \mu)^2.$$

Since
$$E(x_i) = \mu = E(\bar{x}),\ E(x_i - \mu)^2 = \mathrm{var}(x_i) = \sigma^2$$

and
$$E(\bar{x} - \mu)^2 = \mathrm{var}(\bar{x}) = \frac{\sigma^2}{n}$$

$$\therefore \qquad E(s^2) = \sigma^2 - \frac{\sigma^2}{n}$$

$$= \left(\frac{n-1}{n}\right)\sigma^2$$

$$\therefore \qquad E\left(\frac{n}{n-1}s^2\right) = \sigma^2$$

Unbiased estimate of population variance

$$= \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = S^2 \text{ (say)}.$$

15.4-3. Test of Significance of Single Mean

Consider a large random sample with mean $\bar{x}$ from a large population with mean μ and s.d. σ. Then $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$$\therefore \qquad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

The hypothesis to be tested here is the sample has been drawn from a population with mean μ and *s.d.* σ.

The significance is tested with the aid of normal curve and the rules of taking decisions are same as before.

15.4-4. Confidence Limits or Fiducial Limits

Consider a large random sample of size n with mean $\bar{x}$ from a population (not necessarily normal) with mean μ and s.d. σ. Then

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

is nearly a $N(0, 1)$. If σ be known but not μ, there is a range of possible values of μ for which $\bar{x}$ is not significant at any specified level of probability. If $\bar{x}$ is not significant at 5% level of probability, then since

$$P\{|z| > 1\cdot96\} = 0\cdot05,$$

μ must be s.t.

$$\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| < 1\cdot96$$

$$\therefore \qquad \bar{x} - 1\cdot96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1\cdot96\frac{\sigma}{\sqrt{n}}$$

The values $\bar{x} \mp 1\cdot96\frac{\sigma}{\sqrt{n}}$ are called 95% **Fiducial Limits** or **Confidence Limits** *for the mean of the population corresponding to the given sample. The interval $\bar{x}-1\cdot96\frac{\sigma}{\sqrt{n}}$ to $\bar{x}+1\cdot96\frac{\sigma}{\sqrt{n}}$ is called* 95% **Confidence Interval.**

Similarly since $P\{|z| > 2\cdot58\} = 0\cdot01$, 99% Fiducial limits for the population mean are $\bar{x} \mp 2\cdot58\frac{\sigma}{\sqrt{n}}$.

In general if $P\{|z| > z'\} = P'$, $100\,(1-P')\%$. Fiducial limits are $\bar{x} \mp z'\frac{\sigma}{\sqrt{n}}$.

Evidently the limit vary from sample to sample. The totality of values of limits (for given P') for different samples determine the field within which μ is asserted to lie. This field is known as **Confidence Belt.**

Ex. 15-24. *A sample of* 400
Can it be reasonably regarded as
and s.d. 1·3″ ?

Sol. Here $n =$

$\therefore \quad z =$

$\therefore$ The sample may be regar
s.d. 1·3″.

Ex. 15-25. *A sample of* 900
reasonably regarded as a simple
2·61 *cm.*?

Sol. Here $n =$

$\therefore \quad z \simeq$

$\therefore$ The sample can be regard
2·61.

Ex. 15.26. *Mean of* 10 *readi*
of measurements is known to be
length of the rod is 19·9″ ?

Sol. Here $\mu =$

$\therefore \quad z =$

$\therefore$ Sample contradicts the gi

Ex. 15-27. *A sample of* 900 *me*
regarded as a random sample from

Sol. Here $n =$

$\therefore \quad z =$

At 1% level sample cannot be
s.d. 2·3.

Ex. 15-28. *The mean of a cer*
the mean of the samples of 100 *fro*
the sample of 25 *from the distribut*

Sol. Let μ and σ be the mean

Let $\bar{x}$ be the mean of the sam

$\therefore \quad z =$

$\therefore \quad \bar{x} =$

$\therefore \quad P(\bar{x} < 0) =$

$=$

$=$

Ex. 15-29. *A normal populat*
the mean of a simple sample of siz

Sol. Here $\mu =$

$\therefore \quad z =$

$\therefore \quad \bar{x} =$

Ex. 15-24. *A sample of 400 male students is found to have a mean height of 67·47″. Can it be reasonably regarded as a sample from a large population with mean height 67·39″ and s.d. 1·3″ ?*

Sol. Here $n = 400,\ \bar{x} = 67{\cdot}47'',\ \mu = 67{\cdot}39''$ and $\sigma = 1{\cdot}3''$

$\therefore \quad z = 1{\cdot}23\ (< 1{\cdot}96).$

$\therefore$ The sample may be regarded as drawn from the population with mean 67·39″ and *s.d.* 1·3″.

Ex. 15-25. *A sample of 900 members is found to have a mean of 3·4 cm. Can it be reasonably regarded as a simple sample from a large population with mean 3·25 cm s.d. 2·61 cm.?*

Sol. Here $n = 900,\ \bar{x} = 3{\cdot}4,\quad \mu = 3{\cdot}25$ and $\sigma = 2{\cdot}61.$

$\therefore \quad z \simeq 1{\cdot}7\ (< 1{\cdot}96)$

$\therefore$ The sample can be regarded as drawn from a population with mean 3·25 and s.d. 2·61.

Ex. 15.26. *Mean of 10 readings on the length of a given rod is 20″. The s.d. of errors of measurements is known to be 0·1″. Does the result contradict the assumption that the length of the rod is 19·9″ ?*

Sol. Here $\mu = 19{\cdot}9,\ \sigma = 0{\cdot}1,\ n = 10,$ and $\bar{x} = 20,$

$\therefore \quad z = 3{\cdot}162\ (> 3)$

$\therefore$ Sample contradicts the given assumption.

Ex. 15-27. *A sample of 900 members is found to have mean 3·5 cms. Can it be reasonably regarded as a random sample from a large population with mean 3·3 cms and s.d. 2·3 cms ?*

Sol. Here $n = 900,\ \bar{x} = 3{\cdot}5,\ \mu = 3{\cdot}3$ and $\sigma = 2{\cdot}3.$

$\therefore \quad z = 2{\cdot}6\ (> 2{\cdot}58),$

At 1% level sample cannot be regarded as drawn from a population with mean 3·3 and *s.d.* 2·3.

Ex. 15-28. *The mean of a certain normal population is equal to the standard error of the mean of the samples of 100 from that distribution. Find the probability that the mean of the sample of 25 from the distribution will be negative.*

Sol. Let μ and σ be the mean and *s.d.* of the population. Then $\mu = \dfrac{\sigma}{\sqrt{n}} = \dfrac{\sigma}{10}.$

Let $\bar{x}$ be the mean of the sample of size 25.

$$\therefore \quad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{25}} = \frac{\bar{x}-\sigma/10}{\sigma/5} = \frac{5\bar{x}}{\sigma} - \frac{1}{2} \sim N(0, 1)$$

$$\therefore \quad \bar{x} = \frac{\sigma}{5}z + \frac{\sigma}{10}$$

$$\therefore \quad P(\bar{x} < 0) = P\left(\frac{\sigma}{5}z + \frac{\sigma}{10} < 0\right) = P\left(z < -\frac{1}{2}\right)$$

$$= P\left(z > \frac{1}{2}\right) = 0{\cdot}5 - P\left(0 < z < \frac{1}{2}\right)$$

$$= 0{\cdot}5 - (0{\cdot}1915) = 0{\cdot}3085 \qquad \text{(from tables).}$$

Ex. 15-29. *A normal population has mean 0·1 and s.d. 2·1. Find the probability that the mean of a simple sample of size 900 will be negative.*

Sol. Here $\mu = 0{\cdot}1,\ \sigma = 2{\cdot}1,\ n = 900,$

$$\therefore \quad z = \left(\frac{\bar{x}-0{\cdot}1}{2{\cdot}1}\right)30 \sim N(0, 1)$$

$$\therefore \quad \bar{x} = 0{\cdot}07z + 0{\cdot}1$$

$$\therefore \quad P(\bar{x} < 0) = P\left(z < -\frac{10}{7}\right)$$
$$= P(z > 1\cdot43) = 0\cdot5 - P(0 < z < 1\cdot43)$$
$$= 0\cdot5 - 0\cdot4236 = 0\cdot0764.$$

Ex. 15-30. *A research worker wishes to estimate the mean of a population, using a sample sufficiently large, such that the probability will be 0·95 that sample mean will not differ from the true mean by more than 25% of the s.d. How large a sample should be taken ?*

Sol. Let n be the size of the sample.

Now $\quad P\{|\bar{x} - \mu| \leq 0\cdot25\,\sigma\} = 0\cdot95$

$$\therefore \quad P(|z| \leq 0\cdot25\sqrt{n}) = 0\cdot95$$
$$\therefore \quad P(0 < z < 0\cdot25\sqrt{n}) = 0\cdot4750$$
$$\therefore \quad 0\cdot25\sqrt{n} = 1\cdot96$$
$$\therefore \quad n = 62.$$

Ex. 15-31. *If the mean breaking strength of copper wire is 574 lbs with a s.d. of 8·3 lbs, how large a sample must be used in order that there be chance $\frac{1}{100}$ that the mean breaking strength of the wire is less than 571 lbs ?*

Sol. Here $\quad \mu = 574, \ \sigma = 8\cdot3.$

$$\therefore \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \left(\frac{571 - 574}{8\cdot3}\right)\sqrt{n} = -\frac{3}{8\cdot3}\sqrt{n}$$

Now $\quad P\left(z < -\frac{3}{8\cdot3}\sqrt{n}\right) = 0\cdot01$

$$\therefore \quad P\left(|z| < \frac{3}{8\cdot3}\sqrt{n}\right) = 1 - 2\,(0\cdot01) = 0\cdot98$$
$$\therefore \quad P\left(0 < z < \frac{3}{8\cdot3}\sqrt{n}\right) = 0\cdot49$$
$$\therefore \quad \frac{3}{8\cdot3}\sqrt{n} = 2\cdot327 \qquad \text{(from normal tables)}$$
$$\therefore \quad n = \frac{(8\cdot3)^2}{9}(2\cdot327)^2 = 41\cdot46.$$
$$\therefore \quad n \simeq 42.$$

Ex. 15-32. *The guaranteed average life of a certain type of electric light bulbs is 1,000 hours with a s.d. of 125 hours. It is decided to sample the output so as to ensure that 90% of the bulbs do not fall short of the guaranteed average by more than 2·5%. What must be the sample size ?*

Sol. Here $\quad \mu = 1{,}000, \ \sigma = 125$

$$\mu - \bar{x} < \frac{2\cdot5}{100}\mu = 25 \ \textit{i.e.,}\ \bar{x} > \mu - 25 = 975$$
$$\therefore \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{-25}{125/\sqrt{n}} = \frac{-\sqrt{n}}{5}$$
$$\therefore \quad P\left(z > -\frac{\sqrt{n}}{5}\right) = 0\cdot9.$$
$$\therefore \quad P\left(0 < z < \frac{\sqrt{n}}{5}\right) = 0\cdot9 - 0\cdot5 = 0\cdot4$$

$$\therefore \quad \frac{\sqrt{n}}{5} \simeq 1$$
$$\therefore \quad n \simeq 4$$

Ex. 15-33. *It is known that the ... in the universe. It is, however consi... such as to ensure that the mean of th... the true value. How much would be ... for drawing 100 members of a samp...*

Sol. Assume the simple sampli...

Here $\quad \mu = 1$

$$|\bar{x} - \mu| < \frac{0\cdot01}{100}\mu = 0$$
$$\therefore \quad |z| = \Big|$$
$$\therefore \quad \frac{0\cdot01}{10}\sqrt{n} = 3$$
$$\therefore \quad n = 9,$$

$\therefore$ Sampling charges = Rs. ...

Ex. 15-34. *To know the mean w... is taken. The mean weight of this sam... any inference from it about the mean...*

Sol. Here s.d. of the universe is ... the sample is large.

$\therefore$ S.E. of the mean $= \frac{9}{\sqrt{225}}$

Assuming simple sampling conc... probability be s.t.

$$\sqrt{n}\,\Big|$$

i.e., $\quad \bar{x} -$

i.e., $\quad 67 -$

i.e., $\quad 65\cdot2$

Ex. 15-35. *The mean height of ... 2·24" Find the odds against the po... greater than 41·7"*

Sol. $\mu = 41\cdot26, \ \sigma = 2\cdot24, \ n =$

S.E. of the sample mean $= \frac{\sigma}{\sqrt{n}}$

The probability of $\quad z = \frac{\bar{x}}{\sigma}$

is needed.

Since $\quad z \sim N$

$$P(z > 1\cdot96) = 0\cdot$$

$$\therefore \quad \frac{\sqrt{n}}{5} \simeq 1\cdot28 \qquad \text{(from normal tables)}$$

$$\therefore \quad n \simeq 40\cdot96 \simeq 41.$$

Ex. 15-33. *It is known that the mean and s.d. of a variable are respectively* 100 *and* 10 *in the universe. It is, however considered sufficient to draw a sample of sufficient size but such as to ensure that the mean of the sample would be, in all probabilities, within* 0·01% *of the true value. How much would be the cost (exclusive of overhead charges) if the charges for drawing* 100 *members of a sample be one rupee ?*

Sol. Assume the simple sampling conditions hold.

Here $\mu = 100, \; \sigma = 10$

$$|\bar{x} - \mu| < \frac{0\cdot01}{100}\mu = 0\cdot01 \text{ in all probabilities}$$

$$\therefore \quad |z| = \left|\frac{\bar{x}-\mu}{\sigma}\right|\sqrt{n} < \frac{0\cdot01}{10}\sqrt{n} \text{ in all probabilities}$$

$$\therefore \quad \frac{0\cdot01}{10}\sqrt{n} = 3$$

$$\therefore \quad n = 9{,}000{,}000$$

$$\therefore \quad \text{Sampling charges} = \text{Rs. } \frac{9{,}000{,}000}{100} = \text{Rs. } 90{,}000.$$

Ex. 15-34. *To know the mean weight of all* 12*-year old boys in a state, a sample of* 225 *is taken. The mean weight of this sample is found to be* 67 *lbs with a s.d.* 9 *lbs. Can you draw any inference from it about the mean weight of the universe ?*

Sol. Here s.d. of the universe is not given but we can take in its place the sample s.d. as the sample is large.

$$\therefore \quad \text{S.E. of the mean} = \frac{9}{\sqrt{225}} = 0\cdot6.$$

Assuming simple sampling conditions, the mean weight μ of the universe would in all probability be s.t.

$$\sqrt{n}\left|\frac{\bar{x}-\mu}{\sigma}\right| < 3$$

i.e., $\quad \bar{x} - \dfrac{3\sigma}{\sqrt{n}} < \mu < \bar{x} + \dfrac{3\sigma}{\sqrt{n}}$

i.e., $\quad 67 - 1\cdot8 < \mu < 67 + 1\cdot8$

i.e., $\quad 65\cdot2 < \mu < 68\cdot8.$

Ex. 15-35. *The mean height of* 10,000 *children of age* 6 *years is* 41·26″ *and the s.d. is* 2·24″ *Find the odds against the possibility that the mean of a random sample of* 100 *is greater than* 41·7″

Sol. $\mu = 41\cdot26, \; \sigma = 2\cdot24, \; n = 100$

$$\text{S.E. of the sample mean} = \frac{\sigma}{\sqrt{n}} = 0\cdot224.$$

The probability of $\quad z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} > \dfrac{41\cdot7 - 41\cdot26}{0\cdot224} = 1\cdot96.$

is needed.

Since $\quad z \sim N(0, 1).$

$$P(z > 1\cdot96) = 0\cdot5 - P(0 < z < 1\cdot96)$$

$$= 0{\cdot}5 - 0{\cdot}4750 = 0{\cdot}025 = \frac{1}{40}$$

$\therefore$ Odds against are 39 : 1.

Ex. 15-36. *Suppose that the distribution of the statures of men is a normal distribution with s.d. 2·48″. One hundred male students in a large university are measured and their average height is found to be 68·52″. Determine the 98% confidence limits for the mean height of the men of the university.*

Sol. Here $\bar{x} = 68{\cdot}52$, $n = 100$ and $\sigma = 2{\cdot}48$.

$\therefore$ S.E. of sample mean $= \dfrac{\sigma}{\sqrt{n}} = 0{\cdot}248''$.

$$\therefore \qquad z = \frac{68{\cdot}52 - \mu}{0{\cdot}248}.$$

Now since $P\{|z| < 2{\cdot}33\} = 0{\cdot}98$, (from normal tables) μ is needed s.t.

$$\left|\frac{68{\cdot}52 - \mu}{0{\cdot}248}\right| < 2{\cdot}33$$

i.e., $$67{\cdot}9 < \mu < 69{\cdot}1.$$

$\therefore$ 98% confidence limits for μ are 67·9″ and 69·1″.

Ex. 15-37. *The data concerning height measurement for a random sample of individuals from a given population are as follows :*

$$\textit{mean} = 172,\ S.D. = 12,\ n = 100$$

If a large number of samples of the same size were selected at random from the given population, what would be the limits of 2% confidence interval for the true mean ?

Sol. The limits of 2% confidence interval for the true mean means the same thing as 98% confidence limits for the true mean.

$\therefore$ Reqd. limits are

$$172 \pm 2{\cdot}33\left(\frac{12}{\sqrt{100}}\right) \textit{ i.e., } 169{\cdot}2 \text{ and } 174{\cdot}8.$$

Ex. 15-38. *An unbiased coin is thrown n times. It is desired that the relative frequency of the appearance of heads should lie between 0·49 and 0·51. Find the smallest value of n that will ensure this result with 90% confidence.*

Sol. Probability of head (tail) in a single toss $= \dfrac{1}{2}$

$\therefore$ S.E. of proportion of head $= \sqrt{\dfrac{1}{2}\cdot\dfrac{1}{2}\cdot\dfrac{1}{n}} = \dfrac{1}{2\sqrt{n}}$.

Let p be the observed proportion of heads and

$$z = \frac{p - 0{\cdot}5}{1/2\sqrt{n}}.$$

Now since $P\{|z| < 1{\cdot}645\} = 0{\cdot}9$, '$n$' is to determined s.t.

$$\left|\frac{p - 0{\cdot}5}{1/2\sqrt{n}}\right| < 1{\cdot}645$$

i.e., $$0{\cdot}5 - \frac{1{\cdot}645}{2\sqrt{n}} < p < 0{\cdot}5 + \frac{1{\cdot}645}{2\sqrt{n}}$$

$$\therefore \qquad 0{\cdot}5 - \frac{1{\cdot}645}{2\sqrt{n}} = 0{\cdot}49$$

and $\quad 0{\cdot}5 + \dfrac{1{\cdot}645}{2\sqrt{n}} = 0{\cdot}$

Subtracting $\quad \dfrac{1{\cdot}645}{\sqrt{n}} = 0{\cdot}$

$\therefore \quad n = 67$

Ex. 15-39. *If p is the observed trials, prove that the 95% fiducial limi are*

$p \pm 1{\cdot}$

Also show that 99% fiducial limi

$(p - p^2)(2{\cdot}58)^2 = n(p$

Sol. S.E. of proportion of succe

Now $\quad z = p-$

Now since $P\{|z| < 1{\cdot}96\} = 0{\cdot}95$, 9

$p - 1{\cdot}96\sqrt{\dfrac{pq}{n}} < p' < p + 1{\cdot}$

Similarly, since $P\{|z| < 2{\cdot}58\} = 0{\cdot}$

$p' = p \pm$

i.e., $\quad n(p' - p)^2 = (2{\cdot}5$

$= (2{\cdot}5$

1. If p_1, p_2 are observed proportion sizes n_1 and n_2, show that 99% proportions of successes in the p

$p_1 - p_2 \pm 2{\cdot}58$

Also find 95% fiducial limits.

2. A sample of 900 members is fou regarded as a simple sample fron cm ?
3. A simple sample of 1000 memb reasonably regarded as a simple and s.d. 2·6 cm ?
4. The standard deviation of a popul size 100(*i*) the sample mean will d

$\frac{1}{40}$

s of men is a normal distribution
iversity are measured and their
confidence limits for the mean

μ is needed s.t.

″.
or a random sample of individuals

)
elected at random from the given
terval for the true mean ?
te mean means the same thing as

desired that the relative frequency
0·51. *Find the smallest value of n*

$\overline{n}$.

l s.t.

and $$0.5+\frac{1.645}{2\sqrt{n}} = 0.51$$

Subtracting $$\frac{1.645}{\sqrt{n}} = 0.02$$

$$\therefore \quad n = 6765.$$

Ex. 15-39. *If p is the observed proportion of success in n independent Bernoullian trials, prove that the 95% fiducial limits for the population proportion p′, for large samples, are*

$$p \pm 1.96\sqrt{\frac{pq}{n}}.$$

Also show that 99% fiducial limits are the roots of quadratic equation

$$(p-p^2)(2.58)^2 = n(p'-p)^2$$

Sol. S.E. of proportion of successes $= \sqrt{\frac{pq}{n}}$

Now $$z = \frac{p-p'}{\sqrt{\frac{pq}{n}}} \sim N(0, 1)$$

Now since $P\{|z|<1.96\} = 0.95$, 95% fiducial limits are given by

$$\left|\frac{p-p'}{\sqrt{\frac{pq}{n}}}\right| < 1.96$$

$$p-1.96\sqrt{\frac{pq}{n}} < p' < p+1.96\sqrt{\frac{pq}{n}}$$

Similarly, since $P\{|z|<2.58\} = 0.99$, 99% fiducial limits are given by

$$p' = p \pm 2.58\sqrt{\frac{pq}{n}}$$

i.e., $$n(p'-p)^2 = (2.58)^2\, pq = (2.58)^2\, p(1-p)$$
$$= (2.58)^2\,(p-p^2).$$

EXERCISES

1. If p_1, p_2 are observed proportion of successes in two independent sets of trials of large sizes n_1 and n_2, show that 99% fiducial limits for the difference $(p'_1 - p'_2)$ of the proportions of successes in the population are
$$p_1 - p_2 \pm 2.58\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$
Also find 95% fiducial limits.
2. A sample of 900 members is found to have a mean of 3·4 cm. Can it be reasonably regarded as a simple sample from a large population with mean 3·2 cm and s.d. 2·3 cm ? **[Ans.** $z = 2.6$, No.]
3. A simple sample of 1000 members is found to have a mean 3·5 cm. Could it be reasonably regarded as a simple sample from a large population with mean 3·2 cm and s.d. 2·6 cm ? **[Ans.** $z = 3.6$, No.]
4. The standard deviation of a population is 2·7″. Find the probability that in samples of size 100(*i*) the sample mean will differ from the population mean by 0·75 or more and

(*ii*) the sample mean will exceed the population mean by 0·75″ or more.

[Ans. 0·0054; 0·0027]

5. A sample of 900 members is found to have a mean of 3·47 cm. Can it be reasonably regarded as a simple sample from a population with mean 3·23 cm and s.d. 2·31 cm?

[Ans. No.]

15.4-5. Test of significance of the difference between the means of two large samples

Let $\bar{x}_1$, $\bar{x}_2$ be the means of two independent samples of sizes n_1 and n_2 (both n_1 and n_2 are large) from two different populations with means μ_1 and μ_2 and s.d. σ_1 and σ_2 respectively. Then

$$x_1 \sim N\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right) \text{ and } \bar{x}_2 \sim N\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$$

$$\therefore \quad \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\therefore \quad z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

The hypothesis to be tested is 'Are population means same *i.e.*, $\mu_1 = \mu_2$'. Assuming this hypothesis the test statistic becomes

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

The significance is tested with aid of normal curve.

Note 1. If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_2} + \dfrac{1}{n_2}}}.$$

Note 2. If σ is not known, then it is to be estimated from the samples. An unbiased estimate of σ^2 based upon two samples is

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

where $S_1^2 = \dfrac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$ and $S_2^2 = \dfrac{1}{n_2 - 1} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2$

S^2 is unbiased because

$$E(S^2) = \frac{1}{n_1 + n_2 - 2} \{(n_1 - 1)\, E(S_1^2) + (n_2 - 1)\, E(S_2^2)\}$$

$$= \frac{1}{n_1 + n_2 - 2} \{(n_1 - 1)\, \sigma^2 + (n_2 - 1)\, \sigma^2\} = \sigma^2$$

$$[\because\ E(S_1^2) = E(S_2^2) = \sigma^2]$$

Since n_1 and n_2 are large, $n_1 - 1 \sim n_1$ and $n_2 - 1 \sim n_2$

$$\therefore \quad S_1^2 \sim s_1^2 \text{ and } S_2^2 \sim s_2^2$$

where $\quad s_1^2 = \dfrac{1}{n_1} \sum\limits_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$ and $s_2^2 = \dfrac{1}{n_2} \sum\limits_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2$

are sample variances.

$$\therefore \qquad S^2 \simeq \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}.$$

Note. 3. If $\sigma_1^2 \neq \sigma_2^2$ and σ_1 and σ_2 are unknown, these are to be estimated from respective samples. Respective estimates are $S_1^2 \sim s_1^2$ and $S_2^2 \sim s_2^2$.

Ex. 15-40. *The means of simple samples of sizes* 1,000 *and* 2,000 *are* 67·5″ *and* 68·0″ *respectively. Can the samples be regarded as drawn from the same population of s.d.* 2·5″ ?

Sol. $n_1 = 1{,}000,\ n_2 = 2{,}000,\ \bar{x}_1 = 67{\cdot}5,\ \bar{x}_2 = 68{\cdot}0$ and $\sigma = 2{\cdot}5$.

$$\therefore \qquad z \simeq 5{\cdot}2\ (>3)$$

$\therefore$ Hypothesis is wrong and hence the samples cannot be regarded as drawn from the same population of *s.d.* 2·5″.

Ex. 15-41. *A simple sample of heights of* 6,400 *Englishmen has a mean of* 67·85″ *and s.d. of* 2·56″, *while a simple sample of heights of* 1,600 *Australians has a mean* 68·55″ *and a s.d. of* 2·52″. *Do the data indicate that Australians are on the average taller than Englishmen ?*

Sol. Here population standard deviations are not known so these are to be estimated from samples. As samples are large, sample standard deviations can be taken as estimates of population standard deviations.

$$\therefore \qquad \sigma_1 = 2{\cdot}56,\ \sigma_2 = 2{\cdot}52,$$

$$\therefore \qquad z \simeq 10\ (>3).$$

$\therefore$ Difference between sample means is significant.

$\therefore$ Englishmen are on the average smaller than Australians.

Ex. 15-42. *A random sample of* 1200 *men from one state gives their mean pay as Rs.* 40 *p.m. with a s.d. of Rs.* 24 *p.m. and a random sample of* 1600 *men from another state gives their mean pay as Rs.* 36 *p.m. with a s.d. of Rs.* 32 *p.m. Discuss whether the mean levels of pay of men from the two states differ.*

Sol. Here $\quad n_1 = 1200,\ \bar{x}_1 = 40,\ \sigma_1 = 24$ and

$\qquad\qquad n_2 = 1600,\ \bar{x}_2 = 36,\ \sigma_2 = 32$

$$\therefore \qquad z = \frac{4}{\sqrt{\dfrac{(24)^2}{1200} + \dfrac{(32)^2}{1600}}} \simeq 3{\cdot}78\ (>3)$$

$\therefore$ Difference between means is significant and hence mean levels of pay in two states differ.

Ex. 15-43. *Mean and standard deviations calculated from the weights in kgm of students of two groups taken from two universities are given below :*

	Mean	*S.D.*	*Sample size*
University A	55	10	400
University B	57	15	100

Test the significance of the difference between the means.

Sol. Here
$$z = \frac{57-55}{\sqrt{\dfrac{(10)^2}{400} + \dfrac{(15)^2}{100}}} \simeq 1{\cdot}2648\ (< 1{\cdot}96)$$

$\therefore$ Difference between the means is due to fluctuation of sampling only.

Ex. 15-44. *A random sample of* 1000 *farms in a certain year gives an average yield of rice* 2000 *lbs per acre with a s.d. of* 192 *lbs. A random sample of* 1000 *farms in the following year gives an average yield of rice* 2100 *lbs per acre with a s.d. of* 224 *lbs. Show that the data are inconsistent with the hypothesis that the average yield in the country as a whole was the same in the two years.*

Here
$$|z| = \frac{2100-2000}{\sqrt{\frac{(192)^2}{1000}+\frac{(224)^2}{1000}}} \simeq 10{\cdot}7\ (>3)$$

$\therefore$ Data is inconsistent with the hypothesis.

Ex. 15-45. *A potential buyer of light bulbs bought* 50 *bulbs each of two brands. Upon testing these bulbs, he found that brand A had a mean life of* 1282 *hours with a s.d. of* 80 *hours whereas B had a mean life of* 1208 *hours with a s.d. of* 94 *hours. Can the buyer be quite certain that the two brands do differ in quality ?*

Sol. Here
$$n_1 = 50,\ \bar{x}_1 = 1282,\ \sigma_1 = 80$$
and
$$n_2 = 50,\ \bar{x}_2 = 1208,\ \sigma_2 = 94.$$
$$\therefore \quad z \simeq 4{\cdot}2\ (>3)$$

$\therefore$ Difference is significant.

Ex. 15-46. *A random sample of* 200 *villages was taken from a certain district and the average population per village was found to be* 485 *with a s.d. of* 50. *Another random sample of* 200 *villages from the same district gave an average population of* 510 *per village with a s.d. of* 40. *Is the difference between the averages of the two samples significant ? Give reasons.*

Sol. Here
$$n_1 = 200,\ \bar{x}_1 = 485,\ \sigma_1 = 50$$
and
$$n_2 = 200,\ \bar{x}_2 = 510,\ \sigma_2 = 40$$
$$\therefore \quad |z| = \frac{510-485}{\sqrt{\frac{(50)^2+(40)^2}{200}}} \simeq 5{\cdot}5\ (>3)$$

$\therefore$ Difference is significant.

Ex. 15-47. *If* 60 *new entrants in a given university are found to have a mean height of* 68·60″ *and* 50 *seniors a mean height of* 69·51″, *is the evidence conclusive that the mean height of the seniors is greater than that of the new entrants ? Assume the s.d. of the height to be* 2·48″.

Sol. Here
$$n_1 = 60,\ \bar{x}_1 = 68{\cdot}6,\ \sigma_1 = 2{\cdot}48$$
and
$$n_2 = 50,\ \bar{x}_2 = 69{\cdot}51,\ \sigma_1 = 2{\cdot}48$$
$$\therefore \quad |z| \simeq \frac{0\cdot 91}{2\cdot 48\sqrt{\frac{1}{60}+\frac{1}{50}}} \simeq 1{\cdot}92\ (<1{\cdot}96)$$

$\therefore$ Difference is insignificant and hence it cannot be said that the mean height of the seniors is greater than that of the new entrants.

Ex. 15-48. *A sample of* 100 *electric bulbs produced at a factory A showed a mean lifetime of* 1190 *hours and a standard deviation of* 90 *hours. A sample of* 75 *bulbs produced at factory B showed a mean lifetime of* 1230 *hours with a standard deviation of* 12 *hours. Is there a significant difference between the mean lifetimes of the two brands of bulbs at* 5% *level of significance ?*

Sol. Here
$$n_1 = 100,\ \bar{x}_1 = 1190,\ \sigma_1 = 90$$

n_2 =

$\therefore$ $|z|$ =

$\therefore$ Difference is significant

Ex. 15-49. *A certain psychol prisoners : (a) first offenders, (b) re*

Population *San si*

(*i*) *First offenders* 5

(*ii*) *Recidivists* 7

Find the 95% confidence limit

Sol. z =

=

$\therefore$ 95% confidence limits a

6·

i.e., $6{\cdot}43-(0{\cdot}48)(1{\cdot}96) < \mu$

i.e., $5{\cdot}4892 < \mu_1-\mu_2 < 7{\cdot}37$

Ex. 15-50. *Two populations h other. Show that in samples of 45 difference of means will in all prob find the probability that the differe*

Sol. Let μ be the common mea Let $\bar{x}_1$ and $\bar{x}_2$ be the sample mean

Then z

Now $|z|$

i.e., $|\bar{x}_1-\bar{x}_2| < \frac{\sigma}{10}$

Now $P\{|\bar{x}_1-\bar{x}_2| > 0\cdot 05\sigma\}$

$$n_2 = 75, \ \bar{x}_2 = 1230, \ \sigma_2 = 12$$

$$\therefore \qquad |z| = \frac{1230 - 1190}{\sqrt{\dfrac{(90)^2}{100} + \dfrac{(12)^2}{75}}} \simeq 4{\cdot}39 \ (> 1{\cdot}96)$$

$\therefore$ Difference is significant at 5% level.

Ex. 15-49. *A certain psychological test was given to two groups (samples) of army prisoners : (a) first offenders, (b) recidivists. The sample statistics were as follows :*

Population	*Sample size*	*Sample mean*	*Sample s.d.*
(*i*) *First offenders*	580	34·45	8·83
(*ii*) *Recidivists*	786	28·02	8·81

Find the 95% confidence limits of the difference of the means for the two populations.

Sol.

$$z = \frac{(34{\cdot}45 - 28{\cdot}02) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(8{\cdot}83)^2}{580} + \dfrac{(8{\cdot}81)^2}{786}}}$$

$$= \frac{6{\cdot}43 - (\mu_1 - \mu_2)}{0{\cdot}48} \simeq N(0, 1)$$

$\therefore$ 95% confidence limits are given by

$$\left| \frac{6{\cdot}43 - (\mu_1 - \mu_2)}{0{\cdot}48} \right| < 1{\cdot}96$$

i.e., $\quad 6{\cdot}43 - (0{\cdot}48)(1{\cdot}96) < \mu_1 - \mu_2 < (0{\cdot}48)(1{\cdot}96) + 6{\cdot}43$

i.e., $\quad 5{\cdot}4892 < \mu_1 - \mu_2 < 7{\cdot}3708.$

Ex. 15-50. *Two populations have the same mean, but the s.d. of one is twice that of the other. Show that in samples of* 4500 *each drawn under simple sampling conditions the difference of means will in all probability not exceed* $(0{\cdot}1)\,\sigma$ *where* σ *is the smaller s.d. and find the probability that the difference exceeds half that amount.*

Sol. Let μ be the common mean and σ, 2σ be the standard deviations of two populations. Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means.

Then

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{4500} + \dfrac{(2\sigma)^2}{4500}}}$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{\dfrac{\sigma}{30}}$$

Now $\quad |z| < 3$

i.e., $\quad |\bar{x}_1 - \bar{x}_2| < \dfrac{\sigma}{10} = 0{\cdot}1\sigma.$

Now

$$\begin{aligned} P\{|\bar{x}_1 - \bar{x}_2| > 0{\cdot}05\sigma\} &= P\{|z| > 1{\cdot}5\} \\ &= 1 - P\{|z| < 1{\cdot}5\} \\ &= 1 - 2P\{0 < z < 1{\cdot}5\} \\ &= 1 - 2\,(0{\cdot}4332) = 0{\cdot}1336. \end{aligned}$$

Ex. 15·51. *In an intelligence test administrated to 60 boys and 100 girls, the following results were obtained :*

	Mean score	*S.D.*
Boys	114	13
Girls	110	11

Assuming the correlation coefficient between the two to be 0·75, test whether the difference between the means is significant.

Sol. S.E. of the difference between the means

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} - 2\frac{r\sigma_1\sigma_2}{\sqrt{n_1 n_2}}}$$

$$= \sqrt{\frac{(13)^2}{60} + \frac{(11)^2}{100} - 2\frac{(0\cdot75)(13)(11)}{\sqrt{100\cdot 60}}}$$

$$\simeq \sqrt{1\cdot2577} \simeq 1\cdot12$$

$$\therefore \qquad |z| = \frac{4}{1\cdot12} \simeq 3\cdot6\ (>3)$$

$\therefore$ Difference is significant.

EXERCISES

1. The data given below gives the mean and s.d. of stature of two groups of boys taken from a certain city :

Sample size	*Sample mean*	*Sample s.d.*
1145	48·6	2·416
654	50·79	2·53

Find whether the difference between the means significant. **[Ans.** Significant]

2. 64 senior boys from college A and 81 senior boys from college B had mean heights of 68·2″ and 67·3″, respectively. If the s.d. for heights of all senior boys is 2·43″, is the difference between the two groups significant ?

[Ans. $z = 2\cdot21$, significant at 5% level]

3. A random sample of 1000 men from northern region gives their mean wage to be Rs. 21·50 per day with a s.d. of Rs. 1·5. A sample of 1500 men from southern region gives a mean wage of Rs. 21·70 per day with a s.d. of Rs. 2. Discuss whether the mean rate of wages varies as between the two regions. **[Ans.** $z = 2\cdot85$]

4. Two random samples of sizes 1000 and 1500 give following values of mean and s.d. :

Sample size	*Sample mean*	*Sample s.d.*
1000	47	28
1500	49	40

Test whether the difference between means is significant. **[Ans.** No]

5. Intelligence test on two groups of boys and girls, give the following results. Examine if the difference between means is significant.

	Sample mean	*Sample s.d.*	*Sample size*
Girls	84	10	121
Boys	81	12	81

[Ans. Not significant]

6. Two samples of bricks, produced at two different works, were tested for transverse strength with the following results :

	Sample siz
1st Sample	300
2nd Sample	200

Is the difference between th

7. Two populations have the s of the other. Show that in sar the difference of the means smaller s.d. and assuming th find the probability that it e

8. Two population have their m that in the samples of size 20 difference of means will, in a What is the probability that th

15.4.6. Test of significance of di samples

Consider two large independ from two populations with standa The hypothesis to be tested is Assuming this hypothesis, the stat

$z =$

for large samples. Now for large s

var (s_1) =

$\therefore$ var $(s_1 - s_2)$ =

$\therefore$ S.E. $(s_1 - s_2)$ =

$\therefore$ For large samples, z =

The significance is tested with

Ex. 15·52. *Random samples data relating to the heights of male*

	Sa m
University A	67
University B	67

(*i*) *Is the difference between t*
(*ii*) *Is the difference between t*

Sol. (*i*) Here n_1 =

and n_2 =

$\therefore$ z =

$\therefore$ There is no significant d

boys and 100 *girls, the following*

S.D.
13
11

wo to be 0·75, *test whether the*

0·75)(13)(11)
√100·60

ature of two groups of boys taken

Sample s.d.
2·416
2·53

gnificant. [**Ans.** Significant]

rom college *B* had mean heights of
s of all senior boys is 2·43″, is the

s. $z = 2{\cdot}21$, significant at 5% level]

gion gives their mean wage to be
of 1500 men from southern region
of Rs. 2. Discuss whether the mean
[**Ans.** $z = 2{\cdot}85$]

following values of mean and s.d. :

Sample s.d.
28
40

nificant. [**Ans.** No]

give the following results. Examine

d. *Sample size*
121
81
[**Ans.** Not significant]

nt works, were tested for transverse

	Sample size	*Sample mean*	*Sample s.d.*
1st Sample	300	990	240
2nd Sample	200	1000	202

Is the difference between the means significant ? [**Ans.** Not significant]

7. Two populations have the same mean, but the standard deviation of one is twice that of the other. Show that in samples of 500 each drawn under simple random conditions, the difference of the means will in all probability not exceed $0{\cdot}3\ \sigma$ where σ is the smaller s.d. and assuming the distribution of the difference of the means to be normal, find the probability that it exceeds half that amount. [**Ans.** 0·1336]

8. Two population have their means equal but s.d. of one is twice that of the other. Show that in the samples of size 2000 from each drawn under simple sampling conditions, the difference of means will, in all probability, not exceed $0{\cdot}15\ \sigma$ where σ is the smaller s.d. What is the probability that the difference will exceed half this amount ? [**Ans.** 0·1336]

15.4.6. Test of significance of difference between the standard deviations of two large samples

Consider two large independent samples of sizes n_1, n_2 and standard deviations s_1, s_2 from two populations with standard deviations σ_1, σ_2 respectively.

The hypothesis to be tested is 'Are population standard deviations same '*i.e.*, $\sigma_1 = \sigma_2$.' Assuming this hypothesis, the statistic

$$z = \frac{s_1 - s_2}{\text{S.E.}(s_1 - s_2)} \sim N(0, 1)$$

for large samples. Now for large samples drawn from normal populations,

$$\text{var}(s_1) = \frac{\sigma_1^2}{2n_1} \text{ and } \text{var}(s_2) = \frac{\sigma_2^2}{2n_2}$$

$$\therefore \quad \text{var}(s_1 - s_2) = \text{var}(s_1) + \text{var}(s_2) = \frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}$$

$$\therefore \quad \text{S.E.}(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

$$\therefore \quad \text{For large samples, } z = \frac{s_1 - s_2}{\sqrt{\dfrac{\sigma_1^2}{2n_1} + \dfrac{\sigma_2^2}{2n_2}}} \sim N(0, 1)$$

The significance is tested with the aid of normal curve.

Ex. 15-52. *Random samples drawn from two universities A and B gave the following data relating to the heights of male students :*

	Sample mean	*Sample s.d.*	*Sample size*
University A	67·42	2·58	1000
University B	67·25	2·50	1200

(*i*) *Is the difference between the means significant?*

(*ii*) *Is the difference between the standard deviations significant ?*

Sol. (*i*) Here $n_1 = 1000,\ \bar{x}_1 = 67{\cdot}42,\ s_1 = 2{\cdot}58$

and $n_2 = 1200,\ \bar{x}_2 = 67{\cdot}25,\ s_1 = 2{\cdot}50$

$$\therefore \quad z = \frac{0{\cdot}17}{\sqrt{\dfrac{(2{\cdot}58)^2}{1000} + \dfrac{(2{\cdot}5)^2}{1200}}} \simeq 1{\cdot}56\ (< 1{\cdot}96)$$

$\therefore$ There is no significant difference between sample means.

(*ii*) $$z = \frac{0\cdot08}{\sqrt{\dfrac{(2\cdot58)^2}{2000} + \dfrac{(2\cdot5)^2}{2400}}} \simeq 1\cdot04\ (< 1\cdot96)$$

$\therefore$ Sample standard deviations are not significantly different.

Ex. 15-53. *The mean yield of two sets of plots and their variability are as given below. Examine (i) whether the difference in the mean yields of the two sets of plots is significant and (ii) whether the difference in the variability in yields is significant.*

	Set of 40 plots	*Set of 60 plots*
Mean yield per plot	1258	1243
S.D. per plot	34	28

Sol. (*i*) $$|z| = \frac{15}{\sqrt{\dfrac{(34)^2}{40} + \dfrac{(28)^2}{60}}} \simeq 2\cdot3\ (< 2\cdot58)$$

Difference between the mean yields is insignificant at 1% level.

(*ii*) $$|z| = \frac{6}{\sqrt{\dfrac{(34)^2}{2(40)} + \dfrac{(28)^2}{2(60)}}} \simeq 1\cdot3\ (< 1\cdot96)$$

Difference is not significant.

EXERCISES

1. Test whether the difference between the standard deviations is significant, given that

	Size	*s.d.*
Sample *A*	1,392	53·84
Sample *B*	630	56·56

[**Ans.** Not significant]

2. Two samples of sizes 1000 and 800 gave the following results :

	Medians	*S.D.*
1st sample	17·5	2·5
2nd sample	18	2·7

Assuming that samples are independent, test whether the two samples may be regarded as drawn from the universes with same standard deviations. [**Ans.** Yes at 1% level]

Ex. 15-54. *Two samples of sizes* 100 *and* 80 *gave the following results*

	Medians	*S.D.*
1*st sample*	85	7
2*nd sample*	100	8

Test whether the difference between the medians is significant.

Sol. Let σ_1, σ_2 be the standard deviations of two samples of sizes n_1 and n_2. Then assuming the samples to be independent.

S.E. (e) of the difference between the medians

$$= (1\cdot25331)\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Here $\sigma_1 = 7,\ \sigma_2 = 8,\ n_1 = 100$ and $n_2 = 80$

$\therefore$ $e \simeq 1\cdot42$

$$|z| = \frac{\text{Difference between medians}}{e} = \frac{15}{1\cdot42} \simeq 10\cdot6\ (> 3)$$

$\therefore$ Difference is highly significant.

Chi-Sc

16.1. ψ^2 **distribution**

Let $x_1, x_2, \ldots, x_n$ be n indep

Then, each one of

is gamma variate with parameter

$\therefore\ \frac{1}{2}(x_1^2 + x_2^2 + \ldots + x_n^2)$ i

$\therefore$ If $\psi^2 = x_1^2 + x_2^2 + \ldots + x_n$

$\therefore$ Distribution of $\frac{\psi^2}{2}$ is

d

i.e., *d*

where $0 \le \psi^2 < \infty$.

This distribution is known as cl

n is called the degrees of freedom

Remark. (1) Normal distribu distribution for $n=1$.

Hereafter chi-square distributi

$\simeq 1{\cdot}04$ (< 1·96)

tly different.
heir variability are as given below.
f the two sets of plots is significant
ls is significant.
Set of 60 plots
1243
28

2·3 (< 2·58)

t at 1% level.

1·3 (< 1·96)

d deviations is significant, given that
s.d.
53·84
56·56 [**Ans.** Not significant]
lowing results :
S.D.
2·5
2·7
ther the two samples may be regarded
d deviations. [**Ans.** Yes at 1% level]
ve the following results
S.D.
7
8
s is significant.
two samples of sizes n_1 and n_2. Then

$+ \frac{\sigma_2^2}{n_2}$

0 and $n_2 = 80$

en medians $= \frac{15}{1{\cdot}42} \simeq 10{\cdot}6$ (> 3)

16

Chi-Square Distribution

16.1. ψ^2 distribution

Let $x_1, x_2, \ldots, x_n$ be n independent standard normal variates.
Then, each one of

$$\frac{1}{2}x_1^2,\ \frac{1}{2}x_2^2, \ldots\ldots, \frac{1}{2}x_n^2$$

is gamma variate with parameter $\frac{1}{2}$.

$$\therefore\ \frac{1}{2}(x_1^2 + x_2^2 + \ldots\ldots + x_n^2) \text{ is a } \gamma\left(\frac{n}{2}\right) \text{ variate.}$$

$$\therefore \text{ If } \psi^2 = x_1^2 + x_2^2 + \ldots\ldots + x_n^2,\ \frac{\psi^2}{2} \text{ is a } \gamma\left(\frac{n}{2}\right) \text{ variate.}$$

$\therefore$ Distribution of $\frac{\psi^2}{2}$ is

$$dP = \frac{1}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{\psi^2}{2}} \left(\frac{\psi^2}{2}\right)^{\frac{n}{2}-1} d\left(\frac{\psi^2}{2}\right)$$

i.e.,

$$dP = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{\psi^2}{2}} (\psi^2)^{\frac{n}{2}-1} d(\psi^2)$$

where $0 \le \psi^2 < \infty$.

This distribution is known as chi-square distribution and ψ^2 is called chi-square variate. n is called the degrees of freedom associated with chi-square distribution.

Remark. (1) Normal distribution can be regarded as a particular case of chi-square distribution for n=1.

Hereafter chi-square distribution will be written as ψ^2-distribution.

(2) $x_1, x_2, \ldots, x_n$ can be represented by a sample point with co-ordinates $(x_1, x_2 \ldots, x_n)$ in Euclidean hyperspace of n dimensions. If these variates are subjected to a linear constraint, that constraint can be considered to represent a hyperplane. Thus, the effect of this constraint is to lower the dimension by one and hence the number of degrees of freedom associated with ψ^2 will be $n-1$.

In general, if there are p independent linear constraints, the number of $d.f.$ is $n-p$.

16.1.1. M.G.F. of ψ^2-distribution

$$M_0(t) = E\{e^{t\psi^2}\}$$

$$= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{t\psi^2} . e^{-\frac{1}{2}\psi^2} (\psi^2)^{\frac{n}{2}-1} d(\psi^2)$$

$$= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-\frac{1}{2}(1-2t)\psi^2} (\psi^2)^{\frac{n}{2}-1} d(\psi^2)$$

Put $\frac{1}{2}(1-2t)\psi^2 = y$

$$d(\psi^2) = \frac{2dy}{1-2t}$$

$$= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-y} \left(\frac{2y}{1-2t}\right)^{\frac{n}{2}-1} \frac{2dy}{1-2t}$$

$$= \frac{1}{(1-2t)^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-y} y^{\frac{n}{2}-1} dy$$

$$= \frac{1}{(1-2t)^{n/2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{n}{2}\right)$$

$$= (1-2t)^{-\frac{n}{2}}$$

which exists only when, $|2t| < 1$.

16.1.2. Moments and β, γ co-efficients

$$M_0(t) = (1-2t)^{-\frac{n}{2}}$$

$$= 1 + \frac{n}{2}(2t) + \frac{\left(\frac{n}{2}\right)\left(\frac{n}{2}+1\right)}{2!}(2t)^2 + \ldots\ldots$$

$+\ldots.+ \left(\frac{n}{2}\right.$

$\therefore \quad \mu_r'(($

$\therefore \quad$ mean

$\mu_2'(0)$

$\therefore \quad \mu_2$

which gives variance.

$\mu_3'(0)$

μ_3

$\mu_4'(0)$

$\mu_4 = \mu_4'(0) - 4$

$= n(n+2)(n$

$= n^4 + 12n^3$

$= 12n^2 + 48$

β, γ Co-efficients

$$\therefore \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{8}{n},$$

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{8}{n}}$$

with co-ordinates $(x_1, x_2 \ldots, x_n)$
re subjected to a linear constraint,
Thus, the effect of this constraint
f degrees of freedom associated

ts, the number of $d.f.$ is $n-p$.

$$+\ldots+\frac{\left(\frac{n}{2}\right)\left(\frac{n}{2}+1\right)\ldots\left(\frac{n}{2}+\overline{r-1}\right)}{r!}(2t)^r+\ldots$$

$$\therefore \quad \mu_r'(0) = \text{co-efficient of } \frac{t^r}{r!}$$

$$= 2^r\left(\frac{n}{2}\right)\left(\frac{n}{2}+1\right)\ldots\left(\frac{n}{2}+\overline{r-1}\right)$$

$$= n(n+2)\ldots(n+2\overline{r-1})$$

or

$$= \frac{2^r\,\Gamma\left(\frac{n}{2}+r\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$\therefore \quad \text{mean} = \mu_1'(0) = n$$

$$\mu_2'(0) = n(n+2)$$

$$\therefore \quad \mu_2 = n(n+2)-n^2 = 2n$$

which gives variance.

$$\mu_3'(0) = n(n+2)(n+4)$$

$$\mu_3 = \mu_3'(0) - 3\mu_2'(0)\mu_1'(0) + 2\{\mu_1'(0)\}^3$$

$$= n(n+2)(n+4) - 3n^2(n+2) + 2n^3$$

$$= n(n^2+6n+8) - 3(n^3+2n^2) + 2n^3$$

$$= 8n$$

$$\mu_4'(0) = n(n+2)(n+4)(n+6)$$

$$\mu_4 = \mu_4'(0) - 4\mu_3'(0)\mu_1'(0) + 6\mu_2'(0)\{\mu_1'(0)\}^2 - 3\{\mu_1'(0)\}^4$$

$$= n(n+2)(n+4)(n+6) - 4n^2(n+2)(n+4) + 6n^3(n+2) - 3n^4$$

$$= n^4 + 12n^3 + 44n^2 + 48n - 4n^2(n^2+6n+8) + 6n^3(n+2) - 3n^4$$

$$= 12n^2 + 48n$$

β, γ Co-efficients

$$\therefore \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{8}{n}, \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{12}{n}$$

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{8}{n}}, \qquad \gamma_2 = \beta_2 - 3 = \frac{12}{n}$$

$$)^2 + \ldots\ldots$$

16.1.3. Cumulative Function and Cumulants

$$
\begin{aligned}
K_0(t) &= \log M_0(t) \\
&= \log (1-2t)^{-\frac{n}{2}} \\
&= -\frac{n}{2} \log (1-2t) \\
&= \frac{n}{2}\left\{2t+\frac{(2t)^2}{2}+\ldots+\frac{(2t)^r}{r}+\ldots\ldots\right\}
\end{aligned}
$$

$\therefore \quad k_1(0) = \text{co-eff. of } t = n$

$$k_r = \text{co-efficient of } \frac{t^r}{r!} = 2^{r-1}(r-1)!\,n, \quad r \geq 2.$$

16.1.4. Mode

The density function is

$$f(\psi^2) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}\psi^2} (\psi^2)^{\frac{n}{2}-1}$$

$$
\begin{aligned}
f'(\psi^2) &= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}\left\{-\frac{1}{2}e^{-\frac{1}{2}\psi^2}(\psi^2)^{\frac{n}{2}-1}+e^{-\frac{1}{2}\psi^2}\left(\frac{n}{2}-1\right)(\psi^2)^{\frac{n}{2}-2}\right\} \\
&= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}\psi^2}(\psi^2)^{\frac{n}{2}-2}\left\{-\frac{\psi^2}{2}+\frac{n}{2}-1\right\}
\end{aligned}
$$

$\therefore \quad f'(\psi^2) = 0 \Rightarrow \psi^2 = n-2,\ 0$

for $\psi^2 = 0$, $f(\psi^2) = 0$ which is minimum value of $f(\psi^2)$.

$\therefore$ for $\psi^2 = n-2$, $f(\psi^2)$ is maximum.

$\therefore$ Mode $= n-2$.

16.1.5. Limiting form of ψ^2 distribution

Let $\qquad z = \dfrac{\psi^2-n}{\sqrt{2n}}$

Then $\qquad M_0(t)_{\text{of } z} = E\left(e^{tz}\right)$

$$= E\left\{e^{t\,\frac{\psi^2-n}{\sqrt{2n}}}\right\}$$

=

=

$\therefore \quad \log\{M_0(t)_{\text{of } z}\} =$

=

=

$\therefore \ M_0(t)_{\text{of } z} \to e^{\frac{1}{2}t^2}$ as $n \to \infty$.

$\therefore$ z and hence ψ^2 tends to normal

16.1.6. Additive Property of ψ^2-variat

Theorem. The sum of any finite nun

Proof. Let $\psi_1^2, \psi_2^2, \ldots, \psi_n^2$ be n in

of freedom respectively.

Then $\qquad M_0(t)_{\text{of } \psi_i^2} = (1$

Let $\qquad \psi^2 = \psi_1^2+\psi_2^2+\ldots$

Then,

$$M_0(t)_{\text{of } \psi^2} = E\left\{e^{t\psi^2}\right\}$$

$$= E\left\{e^{t(\psi_1^2+\ldots+\psi_n^2}\right.$$

$$\ldots\}$$

$$1)!\,n, \quad r \geq 2.$$

$$\frac{n}{2}-1 + e^{-\frac{1}{2}\psi^2}\left(\frac{n}{2}-1\right)(\psi^2)^{\frac{n}{2}-2}\Bigg\}$$

$$-\frac{\psi^2}{2}+\frac{n}{2}-1\Bigg\}$$

$$(\psi^2).$$

$$= e^{-t\sqrt{\frac{n}{2}}}\, E\left\{e^{\frac{t\psi^2}{\sqrt{2n}}}\right\}$$

$$= e^{-t\sqrt{\frac{n}{2}}}\left\{M_0\left(\frac{t}{\sqrt{2n}}\right)\text{of } \psi^2\right\}$$

$$= e^{-t\sqrt{\frac{n}{2}}}\left(1-\sqrt{\frac{2}{n}}\,t\right)^{-\frac{n}{2}}$$

$$\therefore \qquad \log\{M_0(t)_{\text{of } z}\} = -t\sqrt{\frac{n}{2}}-\frac{n}{2}\log\left\{1-\sqrt{\frac{2}{n}}\,t\right\}$$

$$= -\sqrt{\frac{n}{2}}t+\frac{n}{2}\left\{\sqrt{\frac{2}{n}}t+\frac{\left(\sqrt{\frac{2}{n}}t\right)^2}{2}+\frac{\left(\sqrt{\frac{2}{n}}t\right)^3}{3}+\ldots\right\}$$

$$= \frac{1}{2}t^2+O\left(\frac{1}{\sqrt{n}}\right)$$

$$\to \frac{1}{2}t^2 \text{ as } n\to\infty$$

$$\therefore\ M_0(t)_{\text{of } z} \to e^{\frac{1}{2}t^2} \text{ as } n\to\infty.$$

$\therefore$ z and hence ψ^2 tends to normal variate as $n \to \infty$.

16.1.6. Additive Property of ψ^2-variates

Theorem. The sum of any finite number of independent ψ^2-variates is a ψ^2-variate.

Proof. Let $\psi_1^2, \psi_2^2, \ldots \psi_n^2$ be n independent ψ^2-variates with $n_1, n_2 \ldots n_n$ degrees of freedom respectively.

Then $$M_0(t)_{\text{of } \psi_i^2} = (1-2t)^{n_i/2}, i = 1, 2, \ldots, n$$

Let $$\psi^2 = \psi_1^2+\psi_2^2+\ldots\ldots+\psi_n^2$$

Then,

$$M_0(t)_{\text{of } \psi^2} = E\left\{e^{t\psi^2}\right\}$$

$$= E\left\{e^{t(\psi_1^2+\ldots+\psi_n^2)}\right\}$$

$$= E\{e^{t\psi_1^2}\}E\{e^{t\psi_2^2}\}\ldots.E\{e^{t\psi_n^2}\}$$

$$= (1-2t)^{-\frac{n_1}{2}}.(1-2t)^{-\frac{n_2}{2}}\ldots\ldots(1-2t)^{-\frac{n_n}{2}}$$

$$= (1-2t)^{-\left(\frac{n_1+n_2+\ldots+n_n}{2}\right)}$$

which is the m.g.f. of a ψ^2-variate with $(n_1+\ldots\ldots+n_n)$ d.f.

$\therefore$ ψ^2 is a ψ^2-variate with $(n_1+\ldots\ldots.+n_n)$ d.f.

Ex. 16-1. *If ψ_1^2 and ψ_2^2 are two independent ψ^2-variates with n_1 and n_2 d.f. respectively, then ψ_1^2/ψ_2^2 is a $\beta_2\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate.*

Sol. Distributions of ψ_1^2 and ψ_1^2 respectively, are

$$dP = \frac{1}{2^{n_1/2}\Gamma\left(\frac{n_1}{2}\right)}e^{-\frac{\psi_1^2}{2}}(\psi_1^2)^{\frac{n_1}{2}-1}d(\psi_1^2)$$

$$0<\psi_1^2<\infty$$

and

$$dP = \frac{1}{2^{n_2/2}\Gamma\left(\frac{n_2}{2}\right)}e^{-\frac{1}{2}\psi_2^2}(\psi_2^2)^{\frac{n_2}{2}-1}d(\psi_2^2)$$

$$0<\psi_2^2<\infty$$

The joint distribution of ψ_1^2 and ψ_2^2 is

$$dP = \frac{1}{2^{\frac{n_1+n_2}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}e^{-\frac{1}{2}(\psi_1^2+\psi_2^2)}$$

$$(\psi_1^2)^{\frac{n_1}{2}-1}(\psi_2^2)^{\frac{n_2}{1}-1}d\psi_1^2\,d\psi_2^2$$

$$0<\psi_1^2,\psi_2^2<\infty$$

Put
$$x = \frac{\psi_1^2}{\psi_2^2},\quad y=\psi_2^2$$

$\therefore$
$$\psi_1^2 = xy,\quad \psi_2^2 = y$$

$\therefore$ The joint distribution of x a

The range of x and y are from

$\therefore$ Marginal distribution of x i

$$\therefore \quad \frac{\partial(\psi_1^2, \psi_2^2)}{\partial(x, y)} = \begin{vmatrix} \dfrac{\partial \psi_1^2}{\partial x} & \dfrac{\partial \psi_1^2}{\partial y} \\[2ex] \dfrac{\partial \psi_2^2}{\partial x} & \dfrac{\partial \psi_2^2}{\partial y} \end{vmatrix}$$

$$= \begin{vmatrix} y & x \\ 0 & 1 \end{vmatrix}$$

$$= y$$

$\therefore$ The joint distribution of x and y is

$$dP = \frac{1}{2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{1}{2}y(1+x)} (xy)^{\frac{n_1}{2}-1} y^{\frac{n_2}{2}-1} . y \, dx \, dy$$

$$= \frac{1}{2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{1}{2}y(1+x)} x^{\frac{n_1}{2}-1} y^{\frac{n_1+n_2}{2}-1} dx \, dy$$

The range of x and y are from 0 to ∞.

$\therefore$ Marginal distribution of x is

$$\frac{x^{\frac{n_1}{2}-1}}{2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} dx \int_{y=0}^{\infty} e^{-\frac{1}{2}(1+x)y} y^{\frac{n_1+n_2}{2}-1} dy.$$

$$= \frac{x^{\frac{n_1}{2}-1} dx}{2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \left(\frac{2}{1+x}\right)^{\frac{n_1+n_2}{2}} \int_0^{\infty} e^{-u} u^{\frac{n_1+n_2}{2}-1} du$$

where $u = \frac{1}{2}(1+x)y$

$$= \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{x^{\frac{n_1}{2}-1}}{(1+x)^{\frac{n_1+n_2}{2}}} dx$$

$$= \frac{1}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{x^{\frac{n_1}{2}-1}}{(1+x)^{\frac{n_1}{2}+\frac{n_2}{2}}} dx$$

$\Rightarrow$ x is a $\beta_2\left(\dfrac{n_1}{2}, \dfrac{n_2}{2}\right)$ variate.

Ex. 16-2. *If* ψ_1^2 *and* ψ_2^2 *are independent* ψ^2*-variates with* n_1 *and* n_2 *d.f. respectively, show that*

$$\frac{\psi_1^2}{\psi_1^2+\psi_2^2} \text{ and } \psi_1^2+\psi_2^2$$

are independent. Hence find their distributions.

Sol. The joint dist. of ψ_1^2 and ψ_2^2 is

$$dP = \frac{1}{2^{\frac{n_1+n_2}{2}}\,\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{1}{2}(\psi_1^2+\psi_2^2)}$$

$$(\psi_1^2)^{\frac{n_1}{2}-1}(\psi_2^2)^{\frac{n_2}{2}-1}\,d\psi_1^2\,d\psi_2^2$$

$$0 < \psi_1^2, \psi_2^2 < \infty$$

Put
$$x = \frac{\psi_1^2}{\psi_1^2+\psi_2^2} \text{ and } y = \psi_1^2+\psi_2^2$$

$\therefore$
$$\psi_1^2 = xy \quad \text{and} \quad \psi_2^2 = y(1-x)$$

$$\frac{\partial(\psi_1^2,\psi_2^2)}{\partial(x,y)} = \begin{vmatrix} \dfrac{\partial\psi_1^2}{\partial x} & \dfrac{\partial\psi_1^2}{\partial y} \\ \dfrac{\partial\psi_2^2}{\partial x} & \dfrac{\partial\psi_2^2}{\partial y} \end{vmatrix}$$

$$= \begin{vmatrix} y & x \\ -y & 1-x \end{vmatrix} = y$$

$\therefore$ The joint dist. of x and y is

$$dP = \frac{1}{2^{\frac{n_1+n_2}{2}}\,\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{1}{2}y}(xy)^{\frac{n_1}{2}-1}$$

$$\{y(1-x)\}^{\frac{n_2}{2}-1}\,y\,dx\,dy$$

$$= \left\{\frac{1}{2^{\frac{n_1+n_2}{2}}\,\Gamma\left(\frac{n_1+n_2}{2}\right)} e^{-\frac{1}{2}y}\,y^{\frac{n_1+n_2}{2}-1}\,dy\right\} \times$$

$\Rightarrow$ x and y are independent.

and

$\therefore$ x is a $\beta_1\left(\dfrac{n_1}{2}, \dfrac{n_2}{2}\right)$

and y is a ψ^2-variate with $(n_1 +$

Ex. 16-3. *For a* ψ^2*-variate*

μ

Sol. For a ψ^2-variate with

$\therefore$ M.G.F. about mean is giv

$\therefore$ $M_{\overline{\psi}^2}$

$\therefore$ $\log\{M_{\overline{\psi}^2}($

Differentiating w.r.t. 't'

$\dfrac{1}{M_{\overline{\psi}^2}(t)} M'_{\overline{\psi}^2}$

$\Rightarrow$ $(1-2t)M'_{\overline{\psi}^2}$

Differentiating r times w.r.t.

$(1-2t)M'^{r+1}_{\overline{\psi}^2}(t) - 2rM'^{r}_{\overline{\psi}^2}$

$$\left\{\frac{1}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} x^{\frac{n_1}{2}-1} (1-x)^{\frac{n_2}{2}-1} dx\right\}$$

$\Rightarrow$ x and y are independent. Marginal distributions of x and y respectively are

$$\frac{1}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} x^{\frac{n_1}{2}-1} (1-x)^{\frac{n_2}{2}-1} dx$$

and
$$\frac{1}{2^{\frac{n_1+n_2}{2}} \Gamma\left(\frac{n_1+n_2}{2}\right)} e^{-\frac{1}{2}y} y^{\frac{n_1+n_2}{2}-1} dy$$

$\therefore$ x is a $\beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

and y is a ψ^2-variate with (n_1+n_2) d.f.

Ex. 16-3. *For a ψ^2-variate with n d.f. show that*

$$\mu_{r+1} = 2r(\mu_r + n\mu_{r-1}),\ r \geq 1$$

Sol. For a ψ^2-variate with $n.d.f.$

$$\text{mean} = n$$

$\therefore$ M.G.F. about mean is given by

$$\therefore \quad M_{\overline{\psi}^2}(t) = E\{e^{t(\psi^2-n)}\}$$
$$= e^{-nt} M_0(t)$$
$$= e^{-nt}(1-2t)^{-n/2}$$

$$\therefore \quad \log\left\{M_{\overline{\psi}^2}(t)\right\} = -nt - \frac{n}{2}\log(1-2t)$$

Differentiating w.r.t. 't'

$$\frac{1}{M_{\overline{\psi}^2}(t)} M'_{\overline{\psi}^2}(t) = -n + \frac{n}{2}\frac{2}{1-2t}$$
$$= \frac{2nt}{1-2t}$$

$$\Rightarrow \quad (1-2t)M'_{\overline{\psi}^2}(t) = 2ntM_{\overline{\psi}^2}(t)$$

Differentiating r times w.r.t. 't' by Leibnitz's theorem

$$(1-2t)M^{r+1}{}_{\overline{\psi}^2}(t) - 2rM^r{}_{\overline{\psi}^2}(t) = 2n\{tM^r{}_{\overline{\psi}^2}(t) + rM^{r-1}{}_{\overline{\psi}^2}(t)\} \qquad \ldots(1)$$

Now $\qquad \mu_r = \{M_{\overline{\psi}^2}(t)\}_{t=0}$

$\therefore$ Substituting $t = 0$ in (1)

$$\mu_{r+1} - 2r\mu_r = 2nr\mu_{r-1}$$

$$\Rightarrow \qquad \mu_{r+1} = 2r\,\{\mu_r + n\mu_{r-1}\}.$$

16.1.7. Chief Features of the chi-square Probability curve

The eq. of the ψ^2 probability curve with n d.f. is

$$y = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}\psi^2} (\psi^2)^{\frac{n}{2}-1}$$

$$\therefore \qquad \log y = -\frac{1}{2}\psi^2 + \left(\frac{n}{2}-1\right)\log\psi^2 - \log 2^{n/2} - \log\Gamma\left(\frac{n}{2}\right)$$

Differentiating w.r.t. ψ^2

$$\frac{1}{y}\frac{dy}{d\psi^2} = -\frac{1}{2} + \left(\frac{n}{2}-1\right)\frac{1}{\psi^2}$$

$$= \left\{\frac{(n-2)-\psi^2}{2\psi^2}\right\}.$$

Since $\psi^2 > 0, y > 0,$ we have

for $\qquad n = 1,2, \qquad \dfrac{dy}{d\psi^2} < 0$

and for $n > 2$,

$$\frac{dy}{d\psi^2} > 0 \qquad \text{if } 0 < \psi^2 < n-2$$

$$= 0 \qquad \text{if } \psi^2 = n-2$$

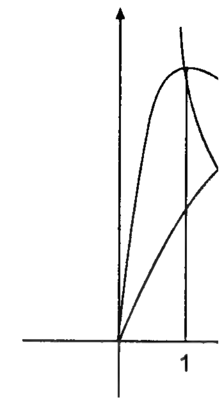
$$< 0 \qquad \text{if } \psi^2 > n-2$$

$\therefore$ For $n = 1, 2$, y decreases continuously as ψ^2 increases and for $n > 2$, y increases or decreases as ψ^2 increases according as $\psi^2 < n-2$ or $\psi^2 > n-2$ and for $\psi^2 = n-2$, $\dfrac{dy}{d\psi^2} = 0$ which implies that y is maximum.

$\therefore$ For all values of n, $y \to 0$ as $\psi^2 \to \infty$.

$\therefore$ ψ^2-axis is an asymptote to the curve.

The shape of curve for $n = 1$,



Ex. 16-4. *If* ψ^2 *is a chi-squ*

normally distributed about mean

Sol. Now $\qquad \sqrt{2\psi}$

if $\qquad 2\psi$

i.e., $\qquad \psi$

i.e., $\qquad \dfrac{\psi^2 -}{\sqrt{2n}}$

i.e., $\qquad \dfrac{\psi^2 -}{\sqrt{2n}}$

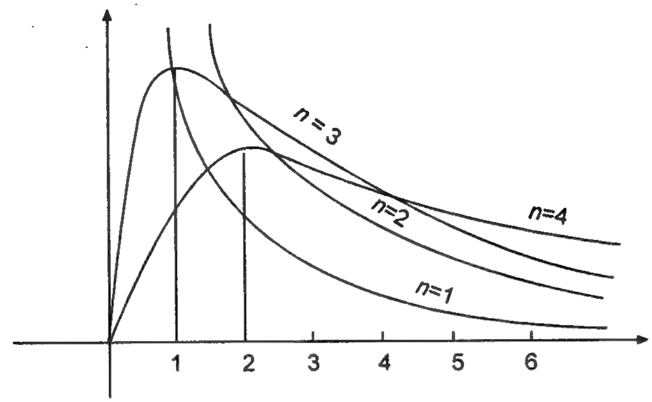
$\therefore \qquad P\left\{\dfrac{\psi^2 - n}{\sqrt{2n}} \le\right.$

Now $\qquad \dfrac{\psi^2 -}{\sqrt{2n}}$

$\therefore \qquad \sqrt{2\psi^2}$

$\Rightarrow \qquad \sqrt{2\psi}$

The shape of curve for $n = 1, 2, 3...$ is shown below :



Ex. 16-4. *If ψ^2 is a chi-square variate with n d.f., show that if n is large $\sqrt{2\psi^2}$ is normally distributed about mean $\sqrt{2n-1}$ with variance unity.*

Sol. Now $$\sqrt{2\psi^2} \le \sqrt{2n-1}+z$$

if $$2\psi^2 \le (2n-1)+z^2+2\sqrt{2n-1}\,z$$

i.e., $$\psi^2 \le n-\frac{1}{2}+\frac{1}{2}z^2+\sqrt{2n-1}\,z$$

i.e., $$\frac{\psi^2-n}{\sqrt{2n}} \le -\frac{1}{2\sqrt{2n}}+\frac{1}{2\sqrt{2n}}z^2+\sqrt{1-\frac{1}{2n}}\cdot z$$

i.e., $$\frac{\psi^2-n}{\sqrt{2n}} \le z, \text{ as } n \text{ is large}$$

$\therefore$ $$P\left\{\frac{\psi^2-n}{\sqrt{2n}} \le z\right\} \approx P\left\{\sqrt{2\psi^2} \le \sqrt{2n-1}+z\right\}$$

$$= P\left\{\sqrt{2\psi^2}-\sqrt{2n-1} \le z\right\}$$

Now $$\frac{\psi^2-n}{\sqrt{2n}} \sim N(0, 1)$$

$\therefore$ $$\sqrt{2\psi^2}-\sqrt{2n-1} \sim N(0,1)$$

$\Rightarrow$ $$\sqrt{2\psi^2} \sim N(\sqrt{2n-1},\ 1).$$

EXERCISES

1. If ψ^2 is a chi-square variate with n d.f., show that if n is large $\sqrt{2\psi^2}$ is normally distributed about mean $\sqrt{2n}$ with variance unity.
2. Show that if x has the standard normal distribution, then x^2 has the chi-square distribution with one degree of freedom.
3. Show that, for 2 degrees of freedom, the probability P of a value of ψ^2 greater than $\psi_0{}^2$ is $\exp.\left(-\frac{1}{2}\psi_0{}^2\right)$ and hence show that

$$\psi_0^2 = 2\log_e\left(\frac{1}{P}\right)$$

4. Prove that for a random sample of size 10 from a normal population with variance 1,

$$P\left\{\sum_{i=1}^{10}(x_i-\bar{x})^2 \geq 25\right\} < 0{\cdot}005$$

(Given that $P(\psi^2(9) < 23{\cdot}59) = 0{\cdot}995$).

16.2. ψ^2-tests

Tests of significance based on ψ^2-distribution are called ψ^2-tests.

Cells. *When a given data is arranged in compartments, the compartments are called cells and the corresponding frequency is called* **Cell Frequency.**

Linear Constraints. *Constraints which involve linear equations in the cell frequencies (i.e., equations containing no squares or higher powers of the frequencies) are called linear constraints.*

Degrees of Freedom. *It is the greatest number of cell frequencies which can be assigned arbitrarily. It is given by*

$$\nu = n - k$$

where n is the total number of cells and k the number of independent constraints.

Definition of ψ^2. *If O_i and e_i be the observed and expected frequencies, the variate ψ^2 is defined by*

$$\psi^2 = \sum_i \frac{(O_i - e_i)^2}{e_i}$$

This variate follows ψ^2-distribution as seen below :

Let there be a random sample of size n whose members are distributed at random in k cells.

Let p_1 = prob. that a member is in ith cell. Then, the prob. that 0_1 members are in 1st cell, 0_2 members in 2nd cell etc., is given by

$$P = \frac{n!}{0_1!0_2!\ldots 0_k!}p_1^{0_1}p_2^{0_2}\ldots\ldots p_k^{0_k}$$

Also $0_1 + 0_2 + \ldots + 0_k = n$

If n is sufficiently large so

factorials can be used.

$\therefore \quad P \approx \prod_{i=1}^{k}$

$= \frac{}{(2\pi}$

$= c\prod_{i=}^{k}$

where $c = \frac{}{(2\pi}$

$\therefore \quad \log P = \log$

Now $e_i = $ exp

$= np_i$

Let $\xi_i = \frac{(0_i}{\sqrt{}}$

$\therefore \quad 0_i = \sqrt{e_i}$

$\therefore \quad \log P \approx \log$

$\therefore \quad \log\frac{P}{c} \approx \sum_{i=1}^{k}$

$\approx -\sum_{i=}^{k}$

t if n is large $\sqrt{2\psi^2}$ is normally

tion, then x^2 has the chi-square

y P of a value of ψ^2 greater than

ıormal population with variance 1,

·005

called ψ^2 -tests.

ıents, the compartments are called
·equency.
ear equations in the cell frequencies of the frequencies) are called linear

ll frequencies which can be assigned

f independent constraints.

ıd expected frequencies, the variate·

w :

mbers are distributed at random in k

the prob. that 0_1 members are in 1st

$_1^{0_1} p_2^{0_2} \ldots\ldots p_k^{0_k}$

Also $0_1 + 0_2 + \ldots + 0_k = n$ ...(1)

If n is sufficiently large so that $0_1, 0_2 \ldots\ldots 0_k$ are not small, Stirling's approximation for factorials can be used.

$$\therefore \quad P \approx \frac{\sqrt{2\pi} e^{-n} . n^{n+\frac{1}{2}}}{\prod_{i=1}^{k} \left\{ \sqrt{2\pi} e^{-0_i} 0_i^{0_i + \frac{1}{2}} \right\}} p_1^{0_1} p_2^{0_2} p_k^{0_k}$$

$$= \frac{1}{(2\pi)^{\frac{k-1}{2}}} \frac{1}{n^{\frac{k-1}{2}}} \frac{1}{(p_1 p_2 \ldots p_k)^{\frac{1}{2}}} \left(\frac{np_1}{0_1} \right)^{0_1 + \frac{1}{2}} \cdots \left(\frac{np_k}{0_k} \right)^{0_k + \frac{1}{2}}$$

$$= c \prod_{i=1}^{k} \left(\frac{np_i}{0_i} \right)^{0_i + \frac{1}{2}}$$

where
$$c = \frac{1}{(2\pi)^{\frac{k-1}{2}} n^{\frac{k-1}{2}} (p_1 p_2 \ldots . p_k)^{\frac{1}{2}}}$$

$$\therefore \quad \log P = \log c + \sum_{i=1}^{k} \left(0_i + \frac{1}{2} \right) \log \frac{np_i}{0_i}$$

Now $\quad e_i$ = expected frequency of ith cell

$$= np_i$$

Let
$$\xi_i = \frac{(0_i - e_i)}{\sqrt{e_i}}$$

$$\therefore \quad 0_i = \sqrt{e_i} \xi_i + e_i$$

$$\therefore \quad \log P \approx \log c + \sum_{i=1}^{k} \left\{ e_i + \sqrt{e_i} \xi_i + \frac{1}{2} \right\} \log \left(\frac{e_i}{e_i + \xi_i \sqrt{e_i}} \right)$$

$$\therefore \quad \log \frac{P}{c} \approx \sum_{i=1}^{k} \left\{ e_i + \sqrt{e_i} \xi_i + \frac{1}{2} \right\} \log \left\{ \frac{1}{1 + \frac{\xi_i}{\sqrt{e_i}}} \right\}$$

$$\approx -\sum_{i=1}^{k} \left\{ e_i + \sqrt{e_i} \xi_i + \frac{1}{2} \right\} \log \left\{ 1 + \frac{\xi_i}{\sqrt{e_i}} \right\}$$

If e_i is large, ξ_i will be small as compared to $\sqrt{e_i}$ and hence the expansion of

$\log\left\{1+\dfrac{\xi_i}{\sqrt{e_i}}\right\}$ is valid.

$\therefore$ Assuming e_i large,

$$\log\frac{P}{c} \approx \sum_{i=1}^{k}\left\{e_i+\sqrt{e_i}\,\xi_i+\frac{1}{2}\right\}\left\{\frac{\xi_i}{\sqrt{e_i}}-\frac{1}{2}\frac{\xi_i^2}{e_i}+\ldots\ldots\right\}$$

$$\approx -\sum_{i=1}^{k}\left\{\xi_i\sqrt{e_i}+\frac{1}{2}\xi_i^2+O\left(\frac{1}{\sqrt{e_i}}\right)\right\}$$

Now

$$\sum_{i=1}^{k}\xi_i\sqrt{e_i} = \sum_{i=1}^{k}\{0_i-e_i\}$$

$$= \sum_{i=1}^{k}0_i-\sum_{i=1}^{k}e_i=0 \qquad \ldots(1)$$

$\therefore$ Neglecting small quantities $\sum_{i=1}^{k}O\left(\dfrac{1}{\sqrt{e_i}}\right)$,

$$\log\frac{P}{c} \approx -\frac{1}{2}\sum_{i=1}^{k}\xi_i^2$$

$$P \approx c e^{-\frac{1}{2}\sum_{i=1}^{k}\xi_i^2}$$

$\Rightarrow$ Each ξ_i is distributed as $N(0,1)$.

Also ξ_i's are connected by linear relation (1)

$$\psi^2 = \sum_{i=1}^{k}\frac{(0_i-e_i)^2}{e_i}=\sum_{i=1}^{k}\xi_i^2$$

is distributed as ψ^2-variate with $(k-1)$ d.f.

Conditions for the Application of ψ^2 test

(*i*) *The members of the sample must be independent.*

(*ii*) *Constraints on the cell-frequencies, if any, should be linear.*

(*iii*) *N, the total frequency must be reasonable large. N should be at least 50, however, few the number of cells.*

(*iv*) *No expected or theoretical cell frequency should be less than 5. It is better if it is greater than or equal to 10.*

Note. *If any expected cell frequency is less than 5, then to apply ψ^2 test this cell is to be*

merged with the preceding or s
adding the cell frequencies of c

Rules of Decision

Let $P = P($

For various fixed values of

of ψ_0^2 are tabulated in the fo

variate is used. Thus role of de

Values of ψ^2 at specified

from the table. Generally 5% a

acceptable at the 5% level of sig
of significance.

Alternately, the probability
and if this is not small the hypo

Remarks. (1) If $\psi^2 = 0, 0$

i.e., observed and expected freq

frequencies differ greatly, ψ^2 is

theory and experiment.

(2) Not only small values o
near to unity may also lead to a

(3) ψ^2-test depends only o

of freedom. It does not make

ψ^2-variate does not involve any

test.

(4) An alternate expressior

(5) The value ψ_0^2 is calle

Uses of ψ^2-test

and hence the expansion of

$\ldots(1)$

nt.

uld be linear.

. N should be at least 50, however,

ld be less than 5. It is better if it is

en to apply ψ^2 test this cell is to be

merged with the preceding or succeeding cells so that the new cell frequency (obtained on adding the cell frequencies of cells merged) is more than 5.

Rules of Decision

Let $$P = P(\psi^2 \geq \psi_0^2)$$

For various fixed values of P and for degrees of freedom n ranging from 1 to 30, value of ψ_0^2 are tabulated in the form of ψ^2 table. For $n > 30$, a property that ψ^2 is normal variate is used. Thus role of decision is as below :

Values of ψ^2 at specified levels of significance for given degrees of freedom are seen from the table. Generally 5% and 1% levels are taken. If $\psi^2_{cal} < \psi^2_{0.05}$, the hypothesis is acceptable at the 5% level of significance otherwise non-acceptable. Similarly, for 1% level of significance.

Alternately, the probability P is determined. If this is small, the hypothesis is rejected and if this is not small the hypothesis is accepted.

Remarks. (1) If $\psi^2 = 0, 0_i = e_i \ \forall_i$

i.e., observed and expected frequencies coincide. On the other hand, if observed and expected frequencies differ greatly, ψ^2 is large. Thus ψ^2 gives a measure of correspondence between theory and experiment.

(2) Not only small values of P lead us to suspect the hypothesis but the value of P very near to unity may also lead to a similar result.

(3) ψ^2-test depends only on the set of observed and expected frequencies and on degrees of freedom. It does not make any assumptions regarding the parent population. Since ψ^2-variate does not involve any population parameter, this test is known as **Non-parametric test.**

(4) An alternate expression for ψ^2 is as below :

$$\psi^2 = \sum_i \frac{(0_i - e_i)^2}{e_i}$$

$$= \sum_i \left\{ \frac{0_i^2}{e_i} + e_i - 2 \cdot 0_i \right\}$$

$$= \sum_i \frac{0_i^2}{e_i} + \sum_i e_i - 2\sum_i 0_i$$

$$= \sum_i \frac{0_i^2}{e_i} - \sum_i 0_i. \qquad (\because \Sigma e_i = \Sigma 0_i)$$

(5) The value ψ_0^2 is called critical value.

Uses of ψ^2-test

Some of the uses of the ψ^2-test are :

(*i*) *To test the goodness of fit.*

(*ii*) *To test the independence of attributes.*

(*iii*) *To test for variance of a normal population.*

(*iv*) *To test the homogeneity of several independent estimates of the population variance.*

(*v*) *To test the homogeneity of several independent estimates of population correlation co-efficient.*

(*vi*) *To combine various probabilities obtained from independent experiments to give a single test of significance.*

Note. *Here only* (*i*) *and* (*ii*) *will be considered.*

16.2.1. The Test of Goodness of Fit

One of the principal uses of ψ^2 distribution is to test how well an observed distribution fits a theoretical one. When ψ^2-test is used in this way, it is called the test of "goodness of fit". The expression within inverted commas may be used in two ways. In the first place it may describe the "fit" of observed to the hypothetical data. In the second it may be used, without reference to a hypothesis, merely to test the merits of a particular formula or a particular curve in graduating a set of values or a series of points, *e.g.*, it may be tested how well a binomial distribution or normal distribution or Poisson distribution fits the given data. The calculations in both the cases are exactly on the same lines.

Ex. 16-5. *In experiments on pea-breeding, Mendal got the following frequencies of seeds : 315 round and yellow; 101 wrinkled and yellow; 108 round and green; 32 wrinkled and green. Theory predicts that the frequencies should be in the proportions 9:3:3:1. Test the correspondence between theory and experiment.*

Sol. Total frequency $= 315 + 101 + 108 + 32 = 556$.

$\therefore$ Expected number of round and yellow seeds $= \dfrac{9}{16} 556 \simeq 313.$

Expected number of wrinkled and yellow seeds $= \dfrac{3}{16} 556 \simeq 104.$

Expected number of round and green seeds $= \dfrac{3}{16} 556 \simeq 104.$

Expected number of wrinkled and green seeds $= \dfrac{1}{16} 556 \simeq 35.$

$$\therefore \qquad \psi^2 = \frac{(313-315)^2}{313} + \frac{(101-104)^2}{104} + \frac{(108-104)^2}{104} + \frac{(32-35)^2}{35}$$

$$\simeq 0{\cdot}013 + 0{\cdot}087 + 0{\cdot}154 + 0{\cdot}257 \simeq 0{\cdot}5.$$

Since there are four expected frequencies, number of d.f.

$$= 4 - 1 = 3.$$

From table $\psi^2_{0\cdot05}$ for 3 d.f. $= 7{\cdot}815$

Now $\qquad \psi^2_{cal} < \psi^2_{0\cdot05}$

$\therefore$ The difference between expected and observed frequencies is not significant at 5%

level of significance.

$\therefore$ Experiment is in agreemen

Ex. 16-6. *A genetical law say. the other parent of blood group N N, and that the average numbers c report on an experiment states as parent, 28·4% were found to be of Do the data in the report conform*

Sol. Total freq. = 162.

Observed frequencies are

$\dfrac{28\cdot4}{100} 162 \simeq 46, \dfrac{42}{100}\cdot 162 \simeq$

and expected frequencies are

$\dfrac{1}{4}\cdot 162 \simeq 40\cdot5, \dfrac{2}{4}\cdot 162 \simeq 81$

$\therefore \qquad \psi$

No. of d

Now $\qquad \psi_{0\cdot05}{}^2$ for 2 d.

$\therefore \qquad \psi_{cal}$

$\therefore$ Hypothesis may be correct

Ex. 16-7. *300 digits were chose*

Digit	0	1	2
Freq.	18	32	28

Test the hypothesis that the di which the data were collected.

Sol. On the assumption that dig frequency of each class

$= \dfrac{300}{10} =$

$\therefore \qquad \psi^2 = \dfrac{1}{30}\{(1$

$\simeq 31\cdot3$

No. of *d.f.* $= 10 - 1 = 9$

From tables, $\psi^2_{0\cdot05}$ for 9 d.f.

$\therefore \qquad \psi^2_{cal} > \psi^2_{0\cdot05}$

$\therefore$ Assumption is wrong.

Ex. 16-8. *200 digits were ch digits were :*

Digits	0	1	2
Freq.	18	19	23

Use ψ^2 test to assess the cor

level of significance.

$\therefore$ Experiment is in agreement with the theory.

Ex. 16-6. *A genetical law says that children having one parent of blood group M and the other parent of blood group N will always be one of the three blood groups M, MN and N, and that the average numbers of children in these groups will be in the ratio 1:2:1. The report on an experiment states as follows of 162 children having one M parent and one N parent, 28·4% were found to be of group M, 42% of group MN and the rest of the group N. Do the data in the report conform to the expected genetic ratio 1:2:1?*

Sol. Total freq. = 162.

Observed frequencies are

$$\frac{28\cdot4}{100}162 \simeq 46, \frac{42}{100}\cdot162 \simeq 68 \text{ and } 162-46-68=48$$

and expected frequencies are

$$\frac{1}{4}\cdot162 \simeq 40\cdot5, \frac{2}{4}\cdot162 \simeq 81 \text{ and } 40\cdot5$$

$$\therefore \qquad \psi^2 = \frac{(5\cdot5)^2}{40\cdot5}+\frac{(13)^2}{81}+\frac{(7\cdot5)^2}{40\cdot5}$$

$$\simeq 4\cdot2.$$

$$\text{No. of d.f.} = 3-1=2.$$

Now $\qquad \psi_{0\cdot05}{}^2$ for 2 d.f. $= 5\cdot99$

$$\therefore \qquad \psi_{cal}{}^2 < \psi_{0\cdot05}{}^2$$

$\therefore$ Hypothesis may be correct and hence genetical law appears to be correct.

Ex. 16-7. *300 digits were chosen at random and found to give the following distribution :*

Digit	0	1	2	3	4	5	6	7	8	9
Freq.	18	32	28	34	42	50	17	23	27	29

Test the hypothesis that the digits were distributed in equal numbers in the table from which the data were collected.

Sol. On the assumption that digits are distributed in equal numbers in the table, expected frequency of each class

$$= \frac{300}{10} = 30$$

$$\therefore \qquad \psi^2 = \frac{1}{30}\{(12)^2+2^2+2^2+4^2+(12)^2+(20)^2+(13)^2+7^2+3^2+1^2\}$$

$$\simeq 31\cdot3$$

No. of *d.f.* = 10 – 1 = 9

From tables, $\psi^2_{0\cdot05}$ for 9 d.f. = 16·92

$$\therefore \qquad \psi^2_{cal} > \psi^2_{0\cdot05}$$

$\therefore$ Assumption is wrong.

Ex. 16-8. *200 digits were chosen at random from a set of tables. The frequencies of digits were :*

Digits	0	1	2	3	4	5	6	7	8	9
Freq.	18	19	23	21	16	25	22	20	21	15

Use ψ^2 test to assess the correctness of hypothesis that the digits were distributed in

equal numbers in the table. Given that the values of ψ^2 are respectively 16·9, 18·3 and 19·7 for 9, 10 and 11 degrees of freedom at 5% level of significance.

Sol. Set the hypothesis 'The digits were distributed in equal numbers in the table'. Then expected frequency of each digit

$$= \frac{200}{10} = 20$$

$$\therefore \quad \psi^2 = \frac{1}{20}\{4+1+9+1+16+25+4+0+1+25\}$$

$$= \frac{86}{20} = 4\cdot 3$$

No. of d.f. = 10 − 1 = 9

Now $\psi^2_{0\cdot 05}$ for 9 d.f. = 16·9

$$\therefore \quad \psi^2_{cal} < \psi^2_{0\cdot 05}$$

∴ Data is consistent with the hypothesis and hence the hypothesis may be correct.

Ex. 16-9. *In 120 throws of a single die, the following distribution of faces were obtained :*

Faces	1	2	3	4	5	6	*Total*
Freq.	30	25	18	10	22	15	120

Test whether these results constitute a refutation of the 'equal probability' hypothesis.

Sol. Set the 'equal probability' hypothesis. Then expected frequency of each face

$$= \frac{120}{6} = 20$$

$$\therefore \quad \psi^2 = \frac{1}{20}\{100+25+4+100+4+25\} = 12\cdot 9$$

$$\text{No. of d.f.} = 6-1 = 5$$

$$\therefore \quad \psi^2_{0\cdot 05} = 11\cdot 07$$

$$\therefore \quad \psi^2_{cal} > \psi^2_{0\cdot 05}$$

∴ The hypothesis is wrong.

Ex. 16-10. *The following figures show the distribution of digits in numbers chosen at random from a telephone directory :*

Digit	0	1	2	3	4	5	6	7	8	9
Freq.	1026	1107	997	966	1075	933	1107	972	964	853 = 10,000

Test whether the digits may be taken to occur equally frequently in the directory.

Sol. Set the hypothesis 'Digits occur equally frequently in the directory'. Then expected frequency of each digit

$$= \frac{10{,}000}{10} = 1{,}000$$

$$\therefore \quad \psi^2 = \frac{1}{1000}\{(26)^2+(107)^2+(3)^2+(34)^2+(75)^2+(67)^2$$

$$+(107)^2+(28)^2+(36)^2+(47)^2\} \simeq 39\cdot 142$$

$$\text{No. of } d.f. = 10-1 = 9$$

Now from table, $\psi^2_{0\cdot 05}$ for 9

$$\therefore \quad \psi^2_{cal} > \psi^2_{0\cdot 05}$$

∴ Hypothesis is certainly wr
frequently in the directory.

Ex. 16-11. *In the construction*
from some logarithm tables and th

Digit	0	1	2
Freq.	1439	1441	1461	14

Use ψ^2-test to assess the cor
chance of being chosen.

Sol. Assuming that each digit h
of each digit

$$= 1500$$

$$\therefore \quad \psi^2 = \frac{1}{1500}(6$$

No. of $d.f.$ = 10 − 1 = 9

Now $\psi^2_{0\cdot 05}$ for 9 $d.f.$ = 16·92

$$\therefore \quad \psi^2_{cal} > \psi^2_{0\cdot 05}$$

∴ Hypothesis is wrong.

Ex. 16-12. *Five dice were thr*
thrown are given below :

No. of dice showing 4, 5 or 6	5
Freq.	8

Calculate ψ^2

Sol. The probability of getting

$$= \frac{1}{2}.$$

∴ By B.D., the expected freque
of

$$96\left(\frac{1}{2}+\right.$$

∴ Expected frequencies are
3, 15, 30, 30, 15, 3

Since the border frequencies a
ones. Doing so

Observed freq.	26
Expected freq.	18

$$\therefore \quad \psi^2 = \frac{64}{18}+\frac{25}{30}$$

Ex. 16-13. *Twelve dice were th*

^{2} *are respectively 16·9, 18·3 and*
significance.
a equal numbers in the table'. Then

$+0+1+25\}$

the hypothesis may be correct.
distribution of faces were obtained :

5	6	*Total*
22	15	120

the 'equal probability' hypothesis.
xpected frequency of each face

$25\} = 12\cdot 9$

ution of digits in numbers chosen at

6	7	8	9
107	972	964	853 = 10,000

ally frequently in the directory.
ntly in the directory'. Then expected

$^2+(34)^2+(75)^2+(67)^2$

$47)^2\} \simeq 39\cdot 142$

Now from table, $\psi^2_{0.05}$ for 9 d.f. = $16\cdot 92$

$$\therefore \qquad \psi^2_{cal} > \psi^2_{0.05}$$

$\therefore$ Hypothesis is certainly wrong and hence digits can't be taken to occur equally frequently in the directory.

Ex. 16-11. *In the construction of a table of random numbers, 15,000 digits were taken from some logarithm tables and the numbers of each digit obtained were as follows:*

Digit	0	1	2	3	4	5	6	7	8	9
Freq.	1439	1441	1461	1452	1494	1454	1613	1491	1482	1519

Use ψ^2-test to assess the correctness of the hypothesis that each digit had an equal chance of being chosen.

Sol. Assuming that each digit had an equal chance of being chosen, expected frequency of each digit

$$= 1500$$

$$\therefore \qquad \psi^2 = \frac{1}{1500}(61)^2+(59)^2+(39)^2+(48)^2+6^2+(46)^2+(113)^2$$

$$+9^2+(18)^2+(19)^2 \simeq 17\cdot 8$$

No. of $d.f. = 10-1 = 9$

Now $\psi^2_{0.05}$ for 9 $d.f.$ = 16·92

$$\therefore \qquad \psi^2_{cal} > \psi^2_{0.05}$$

$\therefore$ Hypothesis is wrong.

Ex. 16-12. *Five dice were thrown 96 times and the number of times 4, 5 or 6 was thrown are given below :*

No. of dice showing 4, 5 or 6	5	4	3	2	1	0
Freq.	8	18	35	24	10	1

Calculate ψ^2

Sol. The probability of getting a 4, 5 or 6 in a throw of a single die

$$= \frac{1}{2}.$$

$\therefore$ By B.D., the expected frequencies are the successive terms in the binomial expansion of

$$96\left(\frac{1}{2}+\frac{1}{2}\right)^5$$

$\therefore$ Expected frequencies are

3, 15, 30, 30, 15, 3

Since the border frequencies are small, these are to be combined with the adjacent ones. Doing so

Observed freq.	26	35	24	11
Expected freq.	18	30	30	18

$$\therefore \qquad \psi^2 = \frac{64}{18}+\frac{25}{30}+\frac{36}{30}+\frac{49}{18} \simeq 8\cdot 31$$

Ex. 16-13. *Twelve dice were thrown 4096 times and a throw of 6 was reckoned as a*

success; the observed frequencies are given below :

No. of successes	0	1	2	3	4	5	6	7 *and over*
Freq.	447	1145	1181	796	380	115	24	8

Find the value of ψ^2 on the hypothesis that dice were unbiased and hence show that the data are consistent with the hypothesis so far as the ψ^2-test is concerned.

Sol. On the hypothesis of unbiased dice the theoretical frequencies are the successive terms in the binomial expansion of

$$4096\left(\frac{5}{6}+\frac{1}{6}\right)^{12}$$

as the probability of success with a throw of one die is $\frac{1}{6}$

∴ Expected frequencies are
459; 1102; 1212; 808; 364; 116; 27 and 8.

$$\therefore \quad \psi^2 = \frac{(12)^2}{459}+\frac{(43)^2}{1102}+\frac{(31)^2}{1212}+\frac{(12)^2}{808}+\frac{(16)^2}{364}+\frac{1^2}{116}+\frac{3^2}{27}+\frac{(8-8)^2}{8}$$

$$\simeq 4{\cdot}00$$

No. of $d.f. = 8-1=7$

Now $\psi_{0{\cdot}05}{}^2$ for 7 $d.f. = 14{\cdot}07$

$$\therefore \quad \psi^2_{\text{cal}} < \psi^2_{0{\cdot}05}$$

∴ The data is consistent with the hypothesis.

Ex. 16-14. *A set of 6 similar coins is tossed 640 times with the following result :*

No. of heads	0	1	2	3	4	5	6
Freq.	7	64	140	210	132	75	12

Calculate the binomial frequencies on the assumption that the coins are symmetrical and test the hypothesis.

Sol. On the assumption that coins are unbiased, the expected frequencies are given by the successive terms in the binomial expansion of

$$640\left(\frac{1}{2}+\frac{1}{2}\right)^6 = 10(1+1)^6$$

$$=10\left(1+6+\frac{6.5}{2}+\frac{6.5.4}{3.2}+\frac{6.5.4.3}{4.3.2.1}+\frac{6.5.4.3.2}{5.4.3.2.1}+1\right)$$

∴ Expected frequencies are :

10, 60, 150, 200, 150, 60, 10

$$\therefore \quad \psi^2 = \frac{3^2}{10}+\frac{4^2}{60}+\frac{(10)^2}{150}+\frac{(10)^2}{200}+\frac{(18)^2}{150}+\frac{(15)^2}{60}+\frac{2^2}{10} \simeq 8{\cdot}6$$

No. of d.f. $= 7-1=6$

Now $\psi^2_{0{\cdot}05}$ for 6 d.f. $= 12{\cdot}59$

$$\therefore \quad \psi^2_{\text{cal}} < \psi^2_{0{\cdot}05}$$

∴ Assumption may be correct.

Ex. 16-15. *12 dice were rollea*
had 5 or 6 on the uppermost face
following table :

No. of dice showing 5 or 6 0 1
Freq. 185 1149 3
9 10
105 14

Fit a binomial dist. and test for

Sol. From the data,

$$\text{A.M.} = \frac{1}{26306}\{1149+6530+1$$

+9

$$=\frac{106602}{26306}$$

Let p be the probability of occu
binomial distribution mean $= np$, est

$$np = \frac{106602}{26306}$$

where n = no. of dice = 12

∴ $p = 0{\cdot}3377$

∴ $q = 1-p = 0{\cdot}$

∴ Expected frequencies are suc
2630

∴ Expected frequencies are :
187, 1146, 3215, 5465, 6269

Since expected frequencies of las
merged.

$$\therefore \quad \psi^2 = \frac{2^2}{187}+\frac{3}{114}$$

$$+\frac{1^2}{133}$$

Now since mean and total freque
frequencies,

No.

Now $\psi_{0{\cdot}05}$

∴

∴ Fit is good.

Ex. 16-16. *The following data s*
during 14 years.

No. of suicides in a state per year 0 1
Freq. 364 376

Ex. 16-15. *12 dice were rolled 26303 times and each time the number of dice which had 5 or 6 on the uppermost face was recorded. The results are given in the form of the following table :*

No. of dice showing 5 or 6	0	1	2	3	4	5	6	7	8
Freq.	185	1149	3265	5475	6114	5194	3067	1331	403
	9	10	11	12					
	105	14	4	—					

Fit a binomial dist. and test for goodness of fit.

Sol. From the data,

$$\text{A.M.} = \frac{1}{26306}\{1149+6530+16425+24456+25970+18402$$
$$+9317+3224+945+140+44\}$$
$$= \frac{106602}{26306}$$

Let p be the probability of occurrence of 5 or 6 in a throw of single die. Then since for binomial distribution mean $= np$, estimate of p is given by

$$np = \frac{106602}{26306}$$

where n = no. of dice = 12

$\therefore \quad p = 0{\cdot}3377$

$\therefore \quad q = 1 - p = 0{\cdot}6623$

$\therefore$ Expected frequencies are successive terms in the binomial expansion of

$$26306\,(0{\cdot}6623 + 0{\cdot}3377)^{12}$$

$\therefore$ Expected frequencies are :

187, 1146, 3215, 5465, 6269, 5115, 3043, 1330, 424, 96, 15, 1, 0.

Since expected frequencies of last two classes are less than 5, last three classes are to be merged.

$$\therefore \quad \psi^2 = \frac{2^2}{187}+\frac{3^2}{1146}+\frac{(50)^2}{3215}+\frac{(10)^2}{5465}+\frac{(155)^2}{6269}+\frac{(79)^2}{5115}+\frac{(24)^2}{3043}$$
$$+\frac{1^2}{1330}+\frac{(21)^2}{424}+\frac{9^2}{96}+\frac{(18-16)^2}{16} \simeq 8{\cdot}201$$

Now since mean and total frequency have been used from the data to obtain expected frequencies,

$$\text{No. of d.f.} = 11 - 2 = 9$$

Now $\quad \psi_{0\cdot05}{}^2$ for 9 d.f. $= 16{\cdot}92$

$$\therefore \quad \psi^2{}_{cal} < \psi^2{}_{0\cdot05}$$

$\therefore$ Fit is good.

Ex. 16-16. *The following data shows the suicides of 1096 women in 8 Punjab cities during 14 years.*

No. of suicides in a state per year	0	1	2	3	4	5	6	7
Freq.	364	376	218	89	33	13	2	1

Fit a Poisson distribution to the data and show that the fit is not good. $(e^{-1.18} = 0\cdot 3075)$.

Sol. The parameter m of the Poisson distribution is to be obtained from the data itself. Since it is equal to the mean of distribution, we have

$$m = \frac{1}{1096}\{0(364) + 1(376) + 2(218) + 3(89) + 4(33) + 5(13) + 6(2) + 7(1) \simeq 1\cdot 18$$

$\therefore$ The theoretical frequencies are

$$1096 . e^{-1.18}\frac{(1\cdot 18)^x}{x!}, x = 0, 1, \ldots\ldots, 7$$

i.e., 337, 398, 235, 92, 27, 6, 1, 0

$$\therefore \qquad \psi^2 = \frac{(27)^2}{337} + \frac{(22)^2}{398} + \frac{(17)^2}{235} + \frac{3^2}{92} + \frac{6^2}{27} + \frac{(16-7)^2}{7} \simeq 17\cdot 6$$

merging last three classes as the expected frequencies of last two classes are less than 5.

Here no. of classes = 6 (as last three classes have been merged)

$\therefore$ No. of d.f. = 6 – 2 = 4 (as mean and total freq. are kept same for expected and observed frequencies).

Now $\qquad \psi^2_{0.05}$ for 4 d.f. $= 9\cdot 49$

$$\therefore \qquad \psi^2_{\text{cal}} > \psi^2_{0.05}$$

$\therefore$ Fit is not good.

Ex. 16-17. *May the following set of observations be regarded as those of a random sample from a Poissonian distribution, given* $e^{-0.5} = \cdot 61$.

Deaths :	0	1	2	3	4	*Total*
Freq. :	122	60	15	2	1	200

Sol. As in Ex. 10-47, theoretical frequencies are

122 61 15 2 0

As expected frequencies of last two cells are less than 5. These cells are to be merged with preceding one

Thus we have

O_i :	122	60	18
e_i :	122	61	17

$$\therefore \qquad \psi^2 = \frac{(122-122)^2}{122} + \frac{(60-61)^2}{61} + \frac{(18-17)^2}{17}$$

$$= \frac{1}{61} + \frac{1}{17} \simeq 0\cdot 08$$

No. of d.f. = 3 – 2 = 1

$$\therefore \qquad \psi^2_{0.05} = 3\cdot 84$$

$$\therefore \qquad \psi^2_{\text{cal}} < \psi^2_{0.05}$$

$\therefore$ Given set of observations can be regarded as those of a random sample from a P.D.

Ex. 16-18. *Fit a normal distribution to the data given below and test the goodness of fit :*

Height (inches)	60-62	63-65	66-68	69-71	72-74
Freq.	5	18	42	27	8

Sol. The A.M. m and *s.d.* σ of

respectively.

The calculations are arranged

Heights	*Class boundaries (X)*	Z
60—62	59·5	– 2·72
63—65	62·5	– 1·70
66—68	65·5	– 0·67
69—71	68·5	0·36
72—74	71·5	1·39
	74·5	2·41

In 2nd column class bound

$Z = \frac{X - 67\cdot 45}{2\cdot 92}$ are written and in

the various values of Z are written. I

obtained by subtracting the success

have the same sign and adding them

in the table above). In 6th column

entries in 5th column by total frequ

$$\therefore \qquad \psi^2 = \frac{(5-4\cdot 1}{4\cdot 13}$$

Since mean, *s.d.* and total frequ

frequencies, number of $d.f. = 5 - 3$

Now $\qquad \psi^2$

$\therefore$

$\therefore$ Fit is good.

1. In a sample of peas from coffee of angular peas is 101. Is this ratio in which they should occu

2. In a Mendalian experiment on pe to occur in the proportion 9 : 3 : observed frequencies were resp results correspond with the theo

Sol. The A.M. m and $s.d.$ σ of the given data can be easily shown to be 67·45″ and 2·92″ respectively.

The calculations are arranged in the table below :

Heights	*Class boundaries (X)*	*Z*	*Areas under normal curve from 0 to Z*	*Areas for each class*	*Expected Freq.*	*Observed Freq.*
60—62	59·5	– 2·72	0·4967	0·0413	4·13	5
63—65	62·5	– 1·70	0·4554	0·2068	20·68	18
66—68	65·5	– 0·67	0·2486 } add	0·3892	38·92	42
69—71	68·5	0·36	0·1406 }	0·2771	27·71	27
72—74	71·5	1·39	0·4177	0·0743	7·43	8
	74·5	2·41	0·4920			

In 2nd column class boundaries (X) are written, in 3rd column the values of $Z = \dfrac{X - 67\cdot 45}{2\cdot 92}$ are written and in 4th column areas under the normal curve from $Z = 0$ to the various values of Z are written. In 5th column areas for each class are written. These are obtained by subtracting the successive areas in the 4th column when the corresponding Z's have the same sign and adding them when Z's have opposite signs (which occurs only once in the table above). In 6th column expected frequencies are written by multiplying the entries in 5th column by total frequency 100.

$$\therefore \qquad \psi^2 = \frac{(5-4\cdot 13)}{4\cdot 13} + \frac{(18-20\cdot 68)^2}{20\cdot 68} + \frac{(42-38\cdot 92)^2}{38\cdot 92} + \frac{(27-27\cdot 71)^2}{27\cdot 71} + \frac{(8-7\cdot 43)^2}{7\cdot 43} \simeq 0\cdot 84$$

Since mean, $s.d.$ and total frequency have been used from the data to obtain expected frequencies, number of $d.f. = 5 - 3 = 2$.

Now $\qquad \psi^2_{0\cdot 05}$ for 2 d.f. $= 5\cdot 99$

$\therefore \qquad \psi^2_{\text{cal}} < \psi^2_{0\cdot 05}$

$\therefore$ Fit is good.

EXERCISES

1. In a sample of peas from coffee plants the number of round peas is 336 and the number of angular peas is 101. Is this in agreement with the Mendalian hypothesis that the ratio in which they should occur is 3 : 1?

   [**Ans.** $\psi^2 = 0\cdot 8$]

2. In a Mendalian experiment on pea-breeding the four possible seed varieties are expected to occur in the proportion 9 : 3 : 3 : 1. In one experiment involving 720 trials the actual observed frequencies were respectively 396, 139, 129 and 56. Examine whether these results correspond with the theory.

   [**Ans.** $\psi^2 = 3\cdot 27$]

3. Genetic theory states that children having one parent of blood type M and the other of blood type N will always be one of the three types *M,MN,N* and the proportions of three types will on average be 1 : 2 : 1. A report states that out of 300 children having one M parent and one N parent, 30% were found to be of type M, 45% type MN and the remainder type N. Test the hypothesis by ψ^2 – test.

[**Ans.** $\psi^2 = 4{\cdot}5$. Hypothesis may be correct]

4. Find the value of ψ^2 for the following table :

Class	A	B	C	D	E
Observed	8	29	44	15	4
Expected freq.	7	24	38	24	7

[**Ans.** 6·8]

5. Find the value of ψ^2 for the following table :

Class	A	B	C	D	E
Observed freq.	8	29	47	16	4
Expected freq.	7	24	38	24	7

[**Ans.** 7·3]

6. The following table gives the number of aircraft accidents that occurred during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Days	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
No. of accidents	14	16	8	12	11	9	14	84

[**Ans.** Uniformly distributed]

7. Five coins are tossed 320 times and the following results are obtained.

No. of heads	0	1	2	3	4	5
Freq.	8	57	110	90	0	5

Test the hypothesis that coins are unbiased.

8. 12 dice were rolled 4,096 times and a throw of 4, 5 or 6 is reckoned as a success, the observed frequencies are given below :

No. of successes	0	1	2	3	4	5	6	7	8
Freq.	0	7	60	198	430	731	948	847	536
	9	10	11	12					
	257	71	11	0					

Apply ψ^2-test to test whether dice can be regarded as unbiased.

[**Ans.** Dice can't be regarded as unbiased]

9. Five dice were thrown 192 times and the number of times 3, 4 or 5 were thrown are given below :

No. of dice throwing 3, 4 or 5	5	4	3	2	1	0
Freq.	6	46	70	48	20	3

Calculate ψ^2 [**Ans.** 16·6]

10. The following is the distribution of 106 eight pig litters according to the number of males in the litters :

No. of males	0	1	2	3	4	5	6	7	8	Total
No. of litters	6	5	8	22	23	25	12	1	4	=106

Fit a binomial distribution unc

goodness of fit. ($\psi^2_{0.05}$ for 8 c

11. Records taken of the number c

No. of male births I

0

1

2

3

4

Test whether the data are cons

and that the chance of a male bi

You may use the table given be

D.F. 1

5% value of ψ^2 3·84

12. One hundred and ninety-two f child being born ic otherwise among the first three children.

No. of albinos 0

No. of families 77

Find the expected frequencies being born an albino and test tl

13. Fit binomial distribution to the

x :	0	1	2	3
f :	3	8	11	15

14. In 1,000 extensive sets of trials the number x of successes are f

x :	0	1	2
f :	305	365	21

Fit a Poisson distribution to the

15. A systematic sample of 100 pag the observed frequency distrib follows :

No. of foreign words per page 0

Freq. 48 2

Graduate the data by a Poisson by ψ^2-test.

16. The table below gives the numbe of 584 pages :

Mistakes per page	0	1
No. of pages	238	20

nt of blood type M and the other of
:s M, MN, N and the proportions of
ites that out of 300 children having
o be of type M, 45% type MN and

- test.

= 4·5. Hypothesis may be correct]

D	E
15	4
24	7

[**Ans.** 6·8]

D	E
16	4
24	7

[**Ans.** 7·3]

accidents that occurred during the
its are uniformly distributed over the

d	Thu	Fri	Sat	Total
	11	9	14	84

[**Ans.** Uniformly distributed]

; results are obtained.

3	4	5
90	0	5

, 5 or 6 is reckoned as a success, the

4	5	6	7	8
430	731	948	847	536

ed as unbiased.

Dice can't be regarded as unbiased]
r of times 3, 4 or 5 were thrown are

3	2	1	0
70	48	20	3

[**Ans.** 16·6]

ig litters according to the number of

5	6	7	8	Total
25	12	1	4	=106

Fit a binomial distribution under the hypothesis that the sex ratio is 1 : 1 and test the goodness of fit. ($\psi^2_{0 \cdot 05}$ for 8 $d.f. = 15 \cdot 51$).

11. Records taken of the number of male and female births in 800 families.

No. of male births	No. of female births	No. of families
0	4	32
1	3	178
2	2	290
3	1	236
4	0	64
		800

Test whether the data are consistent with the hypothesis that the binomial law holds and that the chance of a male birth is equal to that of a female birth, namely $q = p = \dfrac{1}{2}$.

You may use the table given below :

D.F.	1	2	3	4	5
5% value of ψ^2	3·84	5·99	7·82	9·49	11·07

[**Ans.** Binomial law does not hold]

12. One hundred and ninety-two families (for each of which the possibility of an albino child being born is otherwise established) had the following distribution of albinos among the first three children.

No. of albinos	0	1	2	3	Total
No. of families	77	90	20	5	192

Find the expected frequencies on the basis of a theoretical probability 0·25 of a child being born an albino and test the goodness of fit.

[**Ans.** Fit is good]

13. Fit binomial distribution to the following data and test the goodness of fit :

x :	0	1	2	3	4	5	6	7	8	9	Total
f :	3	8	11	15	16	14	12	11	9	1	100

14. In 1,000 extensive sets of trials for an event of small probability the frequencies 'f' of the number x of successes are found to be

x :	0	1	2	3	4	5	6	7
f :	305	365	210	80	28	9	2	1

Fit a Poisson distribution to the data and test the goodness of fit.

15. A systematic sample of 100 pages was taken from the Concise Oxford Dictionary and the observed frequency distribution of foreign words per page was found to be as follows :

No. of foreign words per page	0	1	2	3	4	5	6
Freq.	48	27	12	7	4	1	7

Graduate the data by a Poisson distribution and judge the goodness of your graduation by ψ^2-test.

16. The table below gives the number of mistakes committed per page in typing a manuscript of 584 pages :

Mistakes per page	0	1	2	3	4	5	6	7 and above
No. of pages	238	208	97	30	9	0	2	0

Graduate the data by a Poisson distribution and test the goodness of fit. Present your results in a tabular form.

[Below are given values of ψ^2 with probability P of being exceeded in random sampling; n being the number of degrees of freedom :

$P \rightarrow$ / $n \downarrow$	0·95	0·05	0·01
4	0·71	9·49	13·28
5	1·14	11·07	15·09
6	1·64	12·59	16·81
7	2·17	14·07	18·481

16.2.2. Test of Independence of Attributes

Consider for example the attribute-heights of individuals. Then it may be divided into a large number of parts, *e.g.*, very-tall, tall, medium-sized, short and very short. Thus, the given attribute A can be divided into a number of classes $A_1, A_2, \ldots, A_t$. Similarly any other given attribute B can be divided into classes $B_1, B_2, \ldots, B_s$. Evidently when both attributes A and B are taken into account each one of the classes $A_1, A_2, \ldots, A_t$ would be divided into a large number of subclasses according to $B_1, B_2, \ldots B_s$. Such a classification is called **manifold classification** and a table of the following type is obtained.

Attributes **A**

		A_1	$A_2 \ldots\ldots A_j \ldots\ldots A_t$	Total
	B_1	O_{11}	$O_{12} \ldots\ldots O_{1j} \ldots\ldots O_{1t}$	(B_1)
	B_2	O_{21}	$O_{22} \ldots\ldots O_{2j} \ldots\ldots O_{2t}$	(B_2)
B	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$
	B_i	O_{i1}	$O_{i2} \ldots\ldots O_{ij} \ldots\ldots O_{it}$	(B_i)
	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$
	B_s	O_{s1}	$O_{s2} \ldots\ldots O_{sj} \ldots\ldots O_{st}$	(B_s)
	Total	(A_1)	$(A_2) \ldots\ldots (A_j) \ldots\ldots (A_t)$	N

Such a table is called $s \times t$ **contingency table.** Here N is the total frequency, O_{ij} is the frequency of (i, j)th cell (*i.e.*, a place common to ith row and jth column), $(B_1), (B_2), \ldots\ldots (B_s)$ are totals of rows and $(A_1), (A_2), \ldots\ldots, (A_t)$ are column totals, Evidently

$$N = (A_1) + (A_2) + \ldots + (A_t) = (B_1) + (B_2) + \ldots + (B_s) \qquad \ldots(1)$$

To test whether there is any relationship between A and B, the independence of two attributes is assumed (Null hypothesis). On the basis of this hypothesis expected frequencies of various cells are obtained by keeping the row and the column totals for expected frequencies same as for observed frequencies.

Now proportion of individual.

Since A has no influence on B the classes $(A_1), (A_2), \ldots\ldots, (A_t)$

$\therefore$ Expected number of indivi

Knowing expected frequencie

No. of degrees of freedom as

There are in all $s.t.$ cells. Sinc observed frequencies, there are $(s$ - independent linear constraints

$\therefore$ ν = The no. of d.

This is the number of cells wh

Ex. 16·19. *An opinion poll w reform in 100 members of each of below :*

	Favourable
Party A	40
Party B	42

Test for independence of react 2 d.f. = 5·99).

Sol. Assuming the independen expected frequencies is as below :

	Favourable
Party A	$\frac{82 \times 100}{200} = 41$
Party B	$\frac{82 \times 100}{200} = 41$

$\therefore$ ψ^2

No. of d.f. $= (2-1)(3-1) = 2.$

Now $\psi^2_{0.05}$ for 2 $d.f.$

$\therefore$ ψ

$\therefore$ Hypothesis of independence

Now proportion of individuals belonging to class B_i in the entire data

$$= \frac{(B_i)}{N}.$$

Since A has no influence on B, this proportionality is expected to be maintained in all the classes $(A_1), (A_2), \ldots\ldots\ldots, (A_t)$.

$\therefore$ Expected number of individuals belonging to (i, j)th cell

$$= \frac{(A_j)(B_i)}{N} \qquad i = 1, 2, \ldots\ldots\ldots, s$$

$$j = 1, 2, \ldots\ldots\ldots, t$$

Knowing expected frequencies independence is tested by applying ψ^2-test as usual.

No. of degrees of freedom associated with a s×t contingency table.

There are in all *s.t.* cells. Since row and column totals are kept same for expected and observed frequencies, there are $(s+t)$ constraints. Because of (1) there are only $(s+t-1)$ independent linear constraints

$$\therefore \qquad \nu = \text{The no. of d.f.} = \text{s.t.} - (s+t-1)$$

$$= (s-1)(t-1).$$

This is the number of cells whose frequencies can be arbitrarily assigned.

Ex. 16-19. *An opinion poll was conducted to find the reaction to a proposed civic reform in 100 members of each of the two political parties. The information is tabulated below :*

	Favourable	*Unfavourable*	*Indifferent*
Party A	*40*	*30*	*30*
Party B	*42*	*28*	*30*

Test for independence of reactions with the party affiliations (Given that $\psi^2_{0.05}$ *for* 2 d.f. = 5·99).

Sol. Assuming the independence of reactions with the party affiliations, the table of expected frequencies is as below :

	Favourable	Unfavourable	Indifferent
Party A	$\frac{82 \times 100}{200} = 41$	$\frac{58 \times 100}{200} = 29$	$\frac{60 \times 100}{200} = 30.$
Party B	$\frac{82 \times 100}{200} = 41$	$\frac{58 \times 100}{200} = 29$	$\frac{60 \times 100}{200} = 30.$

$$\therefore \qquad \psi^2 = \frac{1^2}{41} + \frac{1^2}{41} + \frac{1^2}{29} + \frac{1^2}{29} = \frac{2}{41} + \frac{2}{29} = \frac{140}{1189} \simeq 0{\cdot}12.$$

No. of d.f. $= (2-1)(3-1) = 2.$

Now $\qquad \psi^2_{0.05}$ for 2 $d.f.$ = 5·99

$$\therefore \qquad \psi^2_{\text{cal}} < \psi^2_{0.05}$$

$\therefore$ Hypothesis of independence of reactions with the party affiliations may be correct.

Ex. 16-20. *The following table shows the result of inoculation against cholera :*

	Not-attacked	*Attacked*
Inoculated	*431*	*5*
Not-inoculated	*291*	*9*

Is there any significant association between inoculation and attack ? Given that

$$v=1\begin{cases}P=0\cdot074 \text{ for } \psi^2=3\cdot2\\ P=0\cdot069 \text{ for } \psi^2=3\cdot3\end{cases}$$

Sol. Assuming the independence between inoculation and attack, expected frequency table is :

	Not-attacked	Attacked
Inoculated	$\dfrac{722\times436}{736}=427\cdot7$	$\dfrac{14\times436}{736}=8\cdot3$
Not-inoculated	$\dfrac{722\times300}{736}=294\cdot3$	$\dfrac{14\times300}{736}=5\cdot7$

$$\psi^2 = (3\cdot3)^2\left\{\frac{1}{427\cdot7}+\frac{1}{8\cdot3}+\frac{1}{294\cdot3}+\frac{1}{5\cdot7}\right\}=3\cdot28$$

$$v = \text{No. of } d.f. = (2-1)(2-1)=1$$

Now for $\psi^2=3\cdot2$, $\quad P=0\cdot074$

Now for $\psi^2=3\cdot3$, $\quad P=0\cdot069$

Now when ψ^2 increases by 0·1, P decreases by 0·005

$\therefore$ When ψ^2 increases by 0·08, P decreases by $\dfrac{0\cdot005}{0\cdot1}\times0\cdot08$

$$= \cdot0040.$$

$\therefore$ For $\psi^2 = 3\cdot28$, $\quad P = 0\cdot074-0\cdot004=0\cdot07.$

Thus, if the hypothesis is true, the data gave results which would be obtained about 7 times in hundred trials. This is infrequent but not very infrequent. Moreover, the theoretical frequencies in the 'attacked' column are not very large. It will, therefore, be unjustified in rejecting the hypothesis but it can be said that data lead us somewhat to believe that hypothesis is not correct *i.e.*, inoculation and attack are associated.

Ex. 16-21. *From the following table*

		Eye colour in sons	
		Not light	*Light*
Eye colour in fathers	*Not light*	*230*	*148*
	Light	*151*	*471*

test the association between the eye colours of fathers and sons.

Sol. Assuming that there is no association, the expected frequency table is

		Eye colour in sons	
		Not light	Light
Eye colour in fathers	Not light	144	234
	Light	237	385

$\therefore \quad \psi^2 = \dfrac{(230-}{1}$

$\simeq 133.$

No. of d.f. = (2 − 1) (2 − 1) =

$\therefore \quad \psi^2_{0\cdot05} = 3\cdot84$

$\therefore \quad \psi^2_c$

$\therefore$ Assumption is wrong.

Ex. 16-22. *In an experiment* *results were obtained :*

Inoculated
Not inoculated
Examine the effect of vaccine

Sol. Assuming that vaccine h

Inoculated
Not inoculated

$\therefore \quad \psi^2 = \dfrac{(17-1}{17}$

$= 25\left\{\dfrac{1}{17}\right.$

No. of d.f. = (2 − 1)

$\therefore \quad \psi^2_{0\cdot05} = 3\cdot84$

$\therefore \quad \psi^2_{0\cdot05} > \psi^2_{0\cdot05}$

$\therefore$ Assumption is wrong.

Ex. 16-23. *Examine by any s* *voting preference in the election f*

Vote for → Area	*A*
Rural	620
Urban	380
Total	1,000

Sol. Assuming the independe table of expected frequencies is :

Vote for → Area ↓	*A*
Rural	550
Urban	450

$$\therefore \qquad \psi^2 = \frac{(230-144)^2}{144} + \frac{(148-234)^2}{234} + \frac{(151-237)^2}{237} + \frac{(471-385)^2}{385}$$

$$\simeq 133.$$

No. of d.f. $= (2-1)(2-1) = 1$

$$\therefore \qquad \psi^2_{0\cdot05} = 3\cdot84$$

$$\therefore \qquad \psi^2_{\text{cal}} > \psi^2_{0\cdot05}$$

$\therefore$ Assumption is wrong.

Ex. 16-22. *In an experiment on immunization of cattle from tuberculosis the following results were obtained :*

	Affected	*Unaffected*
Inoculated	*12*	*28*
Not inoculated	*13*	*7*

Examine the effect of vaccine in controlling the incidence of the disease.

Sol. Assuming that vaccine has no effect on disease, expected frequency table is,

	Affected	Unaffected
Inoculated	17	23
Not inoculated	8	12

$$\therefore \qquad \psi^2 = \frac{(17-12)^2}{17} + \frac{(23-28)^2}{23} + \frac{(8-13)^2}{8} + \frac{(12-7)^2}{12}$$

$$= 25\left\{\frac{1}{17} + \frac{1}{23} + \frac{1}{8} + \frac{1}{12}\right\} \simeq 7\cdot8.$$

No. of d.f. $= (2-1)(2-1) = 1$

$$\therefore \qquad \psi^2{}_{0\cdot05} = 3\cdot84$$

$$\therefore \qquad \psi^2{}_{0\cdot05} > \psi^2{}_{0\cdot05}$$

$\therefore$ Assumption is wrong.

Ex. 16-23. *Examine by any suitable method, whether the nature of area is related to voting preference in the election for which the data are tabulated below :*

Vote for → Area	*A*	*B*	*Total*
Rural	620	480	1,100
Urban	380	520	900
Total	1,000	1,000	2,000

($\psi_{0\cdot05}{}^2$ *for* 1 *d.f.* $= 3\cdot84$)

Sol. Assuming the independence of voting preference and the nature of the area, the table of expected frequencies is :

Vote for → Area ↓	*A*	*B*
Rural	550	550
Urban	450	450

$$\therefore \qquad \psi^2 = \frac{(70)^2}{550} + \frac{(70)^2}{550} + \frac{(70)^2}{450} + \frac{(70)^2}{450} \simeq 39 \cdot 6 > \psi^2_{0.05} \text{ for 1 d.f.}$$

$\therefore$ Assumption is wrong.

Ex. 16-24. *An investigator into chocolate consumption divided India into eight areas and took a random sample from each, the individuals so obtained being classified as consumers or non-consumers of chocolate. His results were as follows :*

Area number :	1	2	3	4	5	6	7	8	*Total*
Consumers :	56	87	142	71	88	72	100	142	758
Non-consumers :	17	20	58	20	31	23	25	48	242
Total	73	107	200	91	119	95	125	190	1,000

Do these results suggest that the consumption of chocolate varies from place-to-place.

Sol. On the assumption that areas and chocolate consumption are independent *i.e.*, chocolate consumption does not vary from place-to-place, the expected frequency table is

Area number :	1	2	3	4	5	6	7	8
Consumers :	55	81	152	69	90	72	95	144
Non-consumers :	18	26	48	22	29	23	30	46

$$\therefore \qquad \psi^2 = 6 \cdot 28$$

No. of d.f. $= (2-1)(8-1) = 7$

From tables, $\psi^2_{0.05}$ for 7 $d.f. = 14 \cdot 07$

$$\psi^2_{\text{cal}} < \psi^2_{0.05}$$

$\therefore$ Assumption may be correct.

Ex. 16-25. *Deduce that for a $s \times t$ contingency table $\psi^2 \le N(s-1)$ or $\psi^2 \le N(t-1)$ whichever is less.*

Sol. Let e_{ij} be the expected frequency of (i,j)th cell.

$$\text{Then} \qquad e_{ij} = \frac{(A_j)(B_i)}{N}.$$

$$\text{Now} \qquad \psi^2 = \sum_{i=1}^{s}\sum_{j=1}^{t} \frac{(0_{ij} - e_{ij})^2}{e_{ij}} = \sum_i \sum_j \left\{ \frac{0_{ij}{}^2}{e_{ij}} - 20_{ij} + e_{ij} \right\}$$

$$= \sum_i \sum_j \frac{0_{ij}}{e_{ij}} - 2\sum_i \sum_j 0_{ij} + \sum_i \sum_j e_{ij}$$

$$= N \sum_i \sum_j \left\{ \frac{0_{ij}}{(A_j)} \right\} \left\{ \frac{0_{ij}}{(B_i)} \right\} - 2N + N$$

Now since $0_{ij} \le (A_j)$,

$$\frac{0_{ij}}{(A_j)} \le 1$$

$$\therefore \qquad \psi^2 \le N \sum_i \sum_j \frac{0_{ij}}{(B_i)} - N$$

$$= N\left\{\sum_i \frac{\sum_j O_{ij}}{(B_i)} - 1\right\} = N\left\{\sum_i \frac{(B_i)}{(B_i)} - 1\right\}$$

$$\left[\because \sum_j O_{ij} = (B_i)\right]$$

$$= N(s-1).$$

Similarly $\psi^2 \le N(t-1)$,

$$\therefore \qquad \psi^2 \le \min.[N(s-1), N(t-1)].$$

Co-efficient of Contingency

The co-efficient of contingency (C) *is given by*

$$C = \sqrt{\frac{\psi^2}{N+\psi^2}}$$

where $N =$ *total freq.*

Yates Correction of Continuity. *This correction consists in modifying the definition of* ψ^2 *as below :*

$$\psi^2 = \sum \frac{\{|O_i - e_i| - 0\cdot 5\}^2}{e_i}$$

In general, correction is made only when the number of degrees of freedom is v=1. For large samples this yields practically the same results as the uncorrected ψ^2, *but difficulties can arise near critical values.*

Ex. 16-26. *Show that in a 2×2 contingency table where in the frequencies are* $\begin{array}{c|c} a & b \\ \hline c & d \end{array}$, ψ^2 *calculated from the hypothesis of independence is*

$$\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

Sol. Let $\begin{array}{c|c} a' & b' \\ \hline c' & d' \end{array}$ be the expected frequencies obtained on the hypothesis of independence.

Then $$a' = \frac{(a+b)(a+c)}{a+b+c+d}, \quad b' = \frac{(a+b)(b+d)}{a+b+c+d}$$

$$c' = \frac{(a+c)(c+d)}{a+b+c+d} \text{ and } d' = \frac{(b+d)(c+d)}{a+b+c+d}$$

$$\therefore \qquad (a-a')^2 = \left\{a - \frac{(a+b)(a+c)}{a+b+c+d}\right\}^2 = \frac{(ad-bc)^2}{(a+b+c+d)^2}$$

Similarly $(b-b')^2=(c-c')^2=(d-d')^2=\dfrac{(ad-bc)^2}{(a+b+c+d)^2}$

$$\therefore \qquad \psi^2 = \sum\frac{(a-a')^2}{a'} = \frac{(ad-bc)^2}{a+b+c+d}\left\{\frac{1}{(a+b)(a+c)}+\frac{1}{(a+b)(b+d)}\right.$$

$$\left.+\frac{1}{(a+c)(c+d)}+\frac{1}{(b+d)(c+d)}\right\}$$

$$=\frac{(ad-bc)^2}{a+b+c+d}\left\{\frac{a+b+c+d}{(a+b)(a+c)(b+d)}+\frac{a+b+c+d}{(a+c)(b+d)(c+d)}\right\}$$

$$=(ad-bc)^2\left\{\frac{a+b+c+d}{(a+b)(a+c)(b+d)(c+d)}\right\}.$$

Ex. 16-27. *Show that for a 2 × n contingency table,*

$$\psi^2=\sum_r\left\{\frac{N_1N_2\left(\dfrac{\mu_{1r}}{N_1}-\dfrac{\mu_{2r}}{N_2}\right)^2}{\mu_{1r}+\mu_{2r}}\right\}$$

where μ_{1r},μ_{2r} are the two frequencies in the rth column and N_1,N_2 are the marginal sums of the two rows.

Sol. Let η_{1r} and η_{2r} be the expected frequencies in rth column.

Then $$\eta_{1r}=\frac{(\mu_{1r}+\mu_{2r})N_1}{N_1+N_2} \text{ and } \eta_{2r}=\frac{(\mu_{1r}+\mu_{2r}).\,N_2}{N_1+N_2}$$

$$\therefore \qquad \psi^2=\sum_r\left[\frac{\left\{\mu_{1r}-\dfrac{(\mu_{1r}+\mu_{2r})N_1}{N_1+N_2}\right\}^2}{\eta_{1r}}+\frac{\left\{\mu_{2r}-\dfrac{(\mu_{1r}+\mu_{2r})N_2}{N_1+N_2}\right\}^2}{\eta_{2r}}\right]$$

$$=\sum_r\left[\frac{(\mu_{1r}N_2-\mu_{2r}N_1)^2}{(N_1+N_2)(\mu_{1r}+\mu_{2r})}\left\{\frac{1}{N_1}+\frac{1}{N_2}\right\}\right]$$

$$=\sum_r\left\{\frac{N_1N_2\left(\dfrac{\mu_{1r}}{N_1}-\dfrac{\mu_{2r}}{N_2}\right)^2}{\mu_{1r}+\mu_{2r}}\right\}$$

Ex. 16-28. *Show that for ent*

	a_1	a_2
	b_1	b_2
Total	n_1	n_2

$\psi^2=\sum_{i=1}^{r} w_i$

where $p_i=\dfrac{a_i}{n_i}$, p

Sol. Two expected frequenci

Now $q_i=1-p_i$

Ex. 16-29. *In Ex. 16-28 shov*

Ex. 16-28. *Show that for entries in 2 × r contingency table,*

							Total
	a_1	a_2	...	a_i	...	a_r	a
	b_1	b_2	...	b_i	...	b_r	b
Total	n_1	n_2	...	n_i	...	n_r	n

$$\psi^2 = \sum_{i=1}^{r} w_i (p_i - p)^2$$

where $$p_i = \frac{a_i}{n_i},\ p = \frac{a}{n},\ w_i = \frac{n_i}{pq},\ q = \frac{b}{n},\ q_i = 1 - p_i$$

Sol. Two expected frequencies of ith column are $\dfrac{n_i a}{n}$ and $\dfrac{n_i b}{n}$.

$$\therefore \quad \psi^2 = \sum_{i=1}^{r}\left[\frac{n}{n_i a}\left\{a_i - \frac{n_i a}{n}\right\}^2 + \frac{n}{n_i b}\left\{b_i - \frac{n_i b}{n}\right\}^2\right]$$

Now $$q_i = 1 - p_i = 1 - \frac{a_i}{n_i} = \frac{n_i - a_i}{n_i} = \frac{b_i}{n_i} \qquad (\because n_i = a_i + b_i)$$

$$\therefore \quad \psi^2 = \sum_{i=1}^{r}\left[\frac{n_i n}{a}\left\{\frac{a_i}{n_i} - \frac{a}{n}\right\}^2 + \frac{n n_i}{b}\left\{\frac{b_i}{n_i} - \frac{b}{n}\right\}^2\right]$$

$$= \sum_{i=1}^{r}\left[\frac{n_i n}{a}(p_i - p)^2 + \frac{n n_i}{b}(q_i - q)^2\right]$$

$$= \sum_{i=1}^{r}\left[\frac{n n_i}{a}(p_i - p)^2 + \frac{n n_i}{b}\{(1 - p_i) - (1 - p)\}^2\right]$$

$$= \sum_{i=1}^{r} n n_i (p_i - p)^2 \left\{\frac{a+b}{ab}\right\}$$

$$= \sum_{i=1}^{r}\left(\frac{n}{a}\right)\left(\frac{n}{b}\right) n_i (p_i - p)^2 = \sum_{i=1}^{r}\frac{n_i}{pq}(p_i - p)^2$$

$$= \sum_{i=1}^{r} w_i (p_i - p)^2$$

Ex. 16-29. *In Ex. 16-28 show that*

$$\psi^2 = \frac{1}{pq}\left\{\sum_{i=1}^{r}(a_i p_i) - ap\right\}$$

Sol. From Ex. 16-28.

$$\psi^2 = \sum_{i=1}^{r} w_i(p_i - p)^2 = \frac{1}{pq}\sum_{i=1}^{r} n_i\{p_i^2 - 2p_i p + p^2\}$$

$$= \frac{1}{pq}\sum_{i=1}^{r}\{(n_i p_i)p_i - 2(p_i n_i)p + n_i p^2\}$$

$$= \frac{1}{pq}\sum_{i=1}^{r}\{a_i p_i - 2pa_i + n_i p^2\}$$

$$= \frac{1}{pq}\left[\sum_{i=1}^{r}(a_i p_i) - 2p\left(\sum_{i=1}^{r} a_i\right) + p^2\left(\sum_{i=1}^{r} n_i\right)\right]$$

$$= \frac{1}{pq}\left[\sum_{i=1}^{r}(a_i p_i) - 2pa + p^2 n\right]$$

$$= \frac{1}{pq}\left[\sum_{i=1}^{r}(a_i p_i) - 2pa + ap\right]$$

$$= \frac{1}{pq}\left\{\sum_{i=1}^{r}(a_i p_i) - ap\right\}$$

EXERCISES

1. Find the value of ψ^2 for 2×2 contingency table :

Hair colour → Eye colour ↓	Light	Dark
Blue	26	9
Brown	7	18

[**Ans.** 12·6]

2. In a locality 100 persons were randomly selected and asked about their educational achievements. The results are given as below :

		Education		
		Middle	High School	College
Sex	Male	10	15	25
	Female	25	10	15

Can you say that education depends on sex ?

[**Ans.** 9·9, Education depends on sex]

3. From the following table find

		Fair
Sex	Boys	592
	Girls	544

4. In an experiment with immuni were obtained :

Inoculated
Not Inoculated
Examine the effects of vaccine
[**Ans.** Vaccine is ef

5. In an experiment on the immun obtained. Derive your inferenc

Inoculated
Not Inoculated

[**Ans.**

6. In experiments on the Spahling obtained :

Died a

Inoculated
Not inoculated or inoculated with control media
Total

Find the value of ψ^2 and test t

7. The table below gives the data

Inoculated
Not inoculated
Test the effectiveness of inocula

8. Can vaccination be regarded as a following data ?
of 1,482 persons exposed to sma 1,482 persons, 343 were vaccin

9. From the following data test whe economic conditions :

Ex

Economic conditions Good
Not Good

3. From the following table find whether the hair colour and sex are associated :

		Hair Colour				
		Fair	Red	Medium	Dark	Jet black
Sex	Boys	592	119	849	504	36
	Girls	544	97	677	451	14

[**Ans.** Associated]

4. In an experiment with immunization of cattle from tuberculosis, the following results were obtained :

	Affected	Unaffected
Inoculated	12	26
Not Inoculated	16	6

Examine the effects of vaccine in controlling the susceptibility to tuberculosis.

[**Ans.** Vaccine is effective in controlling the susceptibility to tuberculosis]

5. In an experiment on the immunization of goats from anthrox the following results were obtained. Derive your inference on the efficiency of the vaccine :

	Died	Survived
Inoculated	2	10
Not Inoculated	6	6

[**Ans.** Survival is not associated with inoculation of vaccine]

6. In experiments on the Spahlinger anti-tuberculosis vaccine the following results were obtained :

	Died or seriously affected	Unaffected or not seriously affected	Total
Inoculated	6	13	19
Not inoculated or inoculated with control media	8	3	11
Total	14	16	30

Find the value of ψ^2 and test the independence.

[**Ans.** 4·7]

7. The table below gives the data obtained during an epidemic of cholera :

	Attacked	Not attacked
Inoculated	31	469
Not inoculated	185	1,315

Test the effectiveness of inoculation in preventing the attack of cholera.

[**Ans.** Inoculated is effective]

8. Can vaccination be regarded as a preventive measure for small pox as evidenced by the following data ?
of 1,482 persons exposed to small-pox in a locality, 368 in all were attacked. Of these, 1,482 persons, 343 were vaccinated and of these only 35 were attacked.

9. From the following data test whether there is any association between intelligency and economic conditions :

		Intelligency			
		Excellent	Good	Medium	Dull
Economic conditions	Good	48	200	150	80
	Not Good	52	180	190	100

10. A producer of a certain film claimed that his movie was not liked equally by men and women. Accordingly, a sample of men and women was collected. The following are the number of men and women falling into each of the five classes :

	Most liked	More liked	Liked	Not much liked	Disliked
Men	110	591	840	500	30
Women	90	549	670	450	20

Is producer's remark supported by data ?

11. The following table shows the association among 1000 school boys, their general ability and their mathematical ability. Calculate the co-efficient of contingency between the two.

		General ability		
		Good	Fair	Poor
Maths ability	Good	44	22	4
	Fair	265	257	178
	Poor	41	91	98

12. The following data observed for hybrids of Datura :

		Flowers	
		Violet	White
Fruits	Prickly	47	21
	Smooth	12	3

Apply ψ^2 -test-to-test the association between colour of flowers and character of fruits. Given that

$$v = 1 \begin{cases} P = 0\cdot402 \text{ for } \psi^2 = 0\cdot7 \\ P = 0\cdot399 \text{ for } \psi^2 = 0\cdot71 \end{cases}$$

[**Ans.** There is no association]

13. From the following table, test the hypothesis that the flower colour is independent of flatness of leaf :

	Flat leaves	Curled leaves	Total
White flowers	99	36	135
Red flowers	20	5	25

Use the following table giving the values of ψ^2 for 1 d.f. for different values of P.

P :	0·5	0·1	0·05
ψ^2 :	0·455	2·706	3·841

[**Ans.** $\psi^2 = 0{\cdot}5$]

14. Candidates for a degree in Mathematics are required to pass a subsidiary examination in Physics. The table below gives the number of candidates classified according to the grading awarded in two subjects. Test if the performances in the two subjects are independent.

		Class in Maths			
		I	II	Pass	Fail
Class in Physics	I	38	60	50	11
	II	24	70	100	27
	Pass	12	36	91	25

15. Sixteen pieces of photogra[
from nearly white to a ver
sheet and pasted on cards,
scraps from the several she
pack. Twenty observers the
each tint either 'light', 'med
The following table shows t

Name assigned to lower tint		L
	Light	
	Medium	
	Dark	
	Total	2

Show that there is a significa
the name assigned to the oth

16. The following table gives th
nature of work. Test whether t

	Skilled
Male	40
Female	10
Total	50

vie was not liked equally by men and
ıen was collected. The following are
of the five classes :

ked	Not much liked	Disliked
40	500	30
70	450	20

1000 school boys, their general ability
efficient of contingency between the

Fair	Poor
22	4
257	178
91	98

ıra :

Flowers	
olet	White
47	21
12	3

ılour of flowers and character of fruits.

$= 0\cdot7$

$= 0\cdot71$

[**Ans.** There is no association]

ıt the flower colour is independent of

urled leaves	Total
36	135
5	25

for 1 d.f. for different values of P.

0·1	0·05
2·706	3·841

[**Ans.** $\psi^2 = 0\cdot5$]

ıired to pass a subsidiary examination
candidates classified according to the
ɔerformances in the two subjects are

Class in Maths		
II	Pass	Fail
60	50	11
70	100	27
36	91	25

15. Sixteen pieces of photographic paper were printed down to different depths of colour from nearly white to a very deep blackish brown. Small scraps were cut from each sheet and pasted on cards, two scraps on each card one above the other, combining scraps from the several sheets in all possible ways, so that there were 256 cards in the pack. Twenty observers then went through the pack independently, each one naming each tint either 'light', 'medium' or 'dark'.

The following table shows the name assigned to each of the two pieces of paper :

Name assigned to lower tint	Name assigned to upper tint: Light	Medium	Dark	Total
Light	850	571	580	2001
Medium	618	593	455	1666
Dark	540	456	457	1453
Total	2008	1620	1492	5120

Show that there is a significant association between the name assigned to one piece and the name assigned to the other.

16. The following table gives the classification of 100 workers according to sex and the nature of work. Test whether the nature of work is independent of the sex of the worker :

	Skilled	Unskilled	Total
Male	40	20	60
Female	10	30	40
Total	50	50	100

❑❑❑

17

t, F and Z Distributions and Small Sample Tests

17.1. Introduction

Let $x_1, x_2, \ldots\ldots x_n$ be the members of a random sample drawn from a normal population with mean μ and s.d. σ.

Let
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

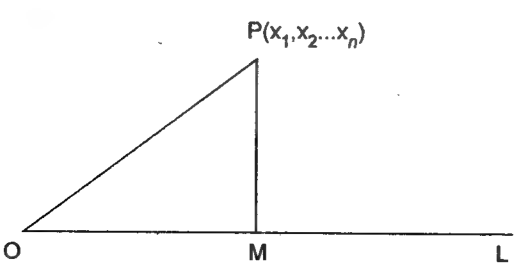
and
$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

The joint distribution of $x_1, x_2, \ldots\ldots x_n$ is

$$dP = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^2 \exp.\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} (x_i-\mu)^2\right\} dx_1, dx_2, \ldots\ldots dx_n$$

Now
$$\sum_{i=1}^{n} (x_i-\mu)^2 = \sum_{i=1}^{n} \{(x_i-\bar{x})+(\bar{x}-\mu)\}^2$$
$$= \sum_{i=1}^{n} \{(x_i-\bar{x})^2+(\bar{x}-\mu)^2+2(\bar{x}-\mu)(x_i-\bar{x})\}$$
$$= \sum_{i=1}^{n} (x_i-\bar{x})^2+n(\bar{x}-\mu)^2 \qquad \left\{\because \sum_{i=1}^{n} (x_i-\bar{x}) = 0\right\}$$
$$= ns^2+n(\bar{x}-\mu)^2$$

Represent the sample values $(x_1, x_2, \ldots\ldots x_n)$ by a pt. P with Co-ordinates $(x_1, x_2, \ldots\ldots x_n)$ in Enclidean hyperspace of n dimension. Let O be the origin. Let OL be the line through O with direction ratios $(1, 1, \ldots\ldots 1)$. Draw $PM \perp OL$.



Let co-ordinates of M be $(\alpha, \alpha, \ldots\ldots \alpha)$ where $(\alpha \neq 0)$

Than d.r.'s of OM are $\alpha, \alpha, \ldots\ldots$

and d.r.'s of PM are $x_1 - \alpha, x$

Since $PM \perp OM$,

$\alpha(x_1 - \alpha) + \alpha(x_2 - \alpha) + \ldots$

$\Rightarrow \quad \alpha =$

$\therefore$ Co-ordinates of M are $(\bar{x}$

$\therefore \quad PM^2 = ($

$\therefore \quad PM =$

and $OM^2 = \bar{x}^2+\ldots+\bar{x}^2 = n$

If $\bar{x}$ and s are kept fixed, P mov

surface of a hypersphere of radius P

$\therefore$ The spherical shell in which

of length $d(OM)$.

$\therefore$ As $\bar{x}$ increases by $d\bar{x}$ and s

$(PM)^{n-2} \,.\, d(PM) \,.\, d(OM)$

$\therefore \quad dP = c$

where c_1 and c_2 are constants.

$\Rightarrow$ s and $\bar{x}$ are independent.

Dist. of $\bar{x}$ is

$dP = c_2$

$\bar{x}$ varies from $-\infty$ to ∞.

$\therefore$ c_2 is given by

Put

Than d.r.'s of OM are $\alpha, \alpha, \ldots\ldots, \alpha$

and d.r.'s of PM are $x_1 - \alpha, x_2 - \alpha, \ldots\ldots x_n - \alpha$.

Since $PM \perp OM$,

$$\alpha(x_1 - \alpha) + \alpha(x_2 - \alpha) + \ldots\ldots + \alpha(x_n - \alpha) = 0$$

$$\Rightarrow \quad \alpha = \frac{x_1 + \ldots\ldots + x_n}{n} = \bar{x}$$

$\therefore$ Co-ordinates of M are $(\bar{x}, \bar{x}, \ldots\ldots \bar{x})$.

$$\therefore \quad PM^2 = (x_1 - \bar{x})^2 + \ldots\ldots + (x_n - \bar{x})^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 = ns^2$$

$$\therefore \quad PM = \sqrt{n}\,.s$$

and $OM^2 = \bar{x}^2 + \ldots\ldots + \bar{x}^2 = n\bar{x}^2 \Rightarrow OM = \sqrt{n}\,\bar{x}$.

If $\bar{x}$ and s are kept fixed, P moves in $(n-1)$ dimensional space orthogonal to OL on the surface of a hypersphere of radius PM and centre M.

$\therefore$ The spherical shell in which P moves has thickness $d(PM)$ and suffers a displacement of length $d(OM)$.

$\therefore$ As $\bar{x}$ increases by $d\bar{x}$ and s by ds, P describes an element of volume proportional to $(PM)^{n-2}\,.\,d(PM)\,.\,d(OM)$

$$= \left\{s\sqrt{n}\right\}^{n-2} \sqrt{n}\, ds . \sqrt{n}\, d\bar{x}$$

$$= \text{constant } s^{n-2}\, ds d\bar{x}$$

$$\therefore \quad dP = \text{const. exp.} \left\{-\frac{1}{2\sigma^2}\{ns^2 + n(\bar{x} - \mu)^2\}\right\} s^{n-2}\, ds d\bar{x}$$

$$= \left\{c_1\, e^{-\frac{1}{2}\frac{ns^2}{\sigma^2}}\, s^{n-2}\, ds\right\} \left\{c_2\, e^{-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\sigma^2/n}}\, d\bar{x}\right\}$$

where c_1 and c_2 are constants.

$\Rightarrow$ s and $\bar{x}$ are independent.

Dist. of $\bar{x}$ is

$$dP = c_2\, e^{-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\sigma^2/n}}\, d\bar{x}$$

$\bar{x}$ varies from $-\infty$ to ∞.

$\therefore$ c_2 is given by

$$c_2 \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\sigma^2/n}}\, d\bar{x} = 1$$

Put $$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = y$$

$$\therefore \quad c_2 \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \frac{\sigma}{\sqrt{n}}\, dy = 1$$

or
$$c_2 \frac{\sigma}{\sqrt{n}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\, dy = 1$$

$$\Rightarrow \qquad c_2 \frac{\sigma}{\sqrt{n}} \cdot \sqrt{2\pi} = 1$$

$$\therefore \qquad c_2 = \sqrt{\frac{n}{2\pi}} \cdot \frac{1}{\sigma}$$

$\therefore$ Dist. of $\bar{x}$ is

$$dP = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma/\sqrt{n}}\, e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2} d\bar{x}$$

$\Rightarrow$ $\bar{x}$ is $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Dist. of s is

$$dP = c_1 e^{-\frac{1}{2}\frac{ns^2}{\sigma^2}} s^{n-2}\, ds$$

s varies from 0 to ∞.

$\therefore$ c_1 is given by

$$c_1 \int_0^{\infty} e^{-\frac{1}{2}\frac{ns^2}{\sigma^2}} s^{n-2}\, ds = 1$$

Put
$$\frac{1}{2}\frac{ns^2}{\sigma^2} = y$$

$$\therefore \qquad s\,ds = \frac{\sigma^2}{n}\, dy$$

$$\therefore \qquad c_1 \cdot \frac{\sigma^2}{n} \int_0^{\infty} e^{-y} \left(\frac{2\sigma^2}{n} y\right)^{\frac{n-3}{2}} dy = 1$$

i.e.,
$$\frac{c_1}{2}\left(\frac{2\sigma^2}{n}\right)^{\frac{n-1}{2}} \int_0^{\infty} e^{-y} . y^{\frac{n-1}{2}-1}\, dy = 1$$

$$\Rightarrow \qquad \frac{c_1}{2}\left(\frac{2\sigma^2}{n}\right)^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) = 1$$

$$\Rightarrow \qquad c_1 = 2\left(\frac{n}{2\sigma^2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)}$$

$\therefore$ Dist. of s is

$$dP = 2\left(\frac{n}{2\sigma^2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{1}{2}\cdot\frac{ns^2}{\sigma^2}} s^{n-2}\, ds$$

$\therefore$ $\frac{ns^2}{\sigma^2}$ is a ψ^2 variate with (n

and $\frac{1}{2}\frac{ns^2}{\sigma^2}$ is a $\gamma\left(\frac{n-1}{2}\right)$ varia

Remark

$E(s^2) =$

Put $\frac{ns^2}{2\sigma^2} = y$

$= \sigma^2$

$\therefore$ $E\left\{\frac{ns^2}{n-1}\right\} = \sigma^2$

$\therefore$ $\frac{ns^2}{n-1}$ is an unbiased estin

17.2. Student's t-Distribution

Student's t statistic is defined by

$$= \frac{1}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}\, e^{-\frac{1}{2}\cdot\frac{ns^2}{\sigma^2}} \left(\frac{ns^2}{\sigma^2}\right)^{\frac{n-3}{2}} d\left(\frac{ns^2}{\sigma^2}\right)$$

$$= \frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\, e^{-\frac{1}{2}\frac{ns^2}{\sigma^2}} \left(\frac{1}{2}\frac{ns^2}{\sigma^2}\right)^{\frac{n-1}{2}-1} d\left(\frac{ns^2}{2\sigma^2}\right)$$

$\therefore$ $\dfrac{ns^2}{\sigma^2}$ is a ψ^2 variate with $(n-1)$ d.f.

and $\dfrac{1}{2}\dfrac{ns^2}{\sigma^2}$ is a $\gamma\left(\dfrac{n-1}{2}\right)$ variate.

Remark

$$E(s^2) = \frac{1}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty s^2 e^{-\frac{1}{2}\frac{ns^2}{\sigma^2}} \left(\frac{ns^2}{\sigma^2}\right)^{\frac{n-3}{2}} d\left(\frac{ns^2}{\sigma^2}\right).$$

Put $\dfrac{ns^2}{2\sigma^2} = y$

$$= \frac{1}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty \frac{2\sigma^2}{n} y.e^{-y}\,(2y)^{\frac{n-3}{2}}\,(2dy)$$

$$= \frac{2}{\Gamma\left(\frac{n-1}{2}\right)}\cdot\frac{\sigma^2}{n} \int_0^\infty e^{-y}\, y^{\frac{n+1}{2}-1}\, dy$$

$$= \frac{2\sigma^2}{n}\cdot\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\,\Gamma\left(\frac{n+1}{2}\right)$$

$$= \frac{2\sigma^2}{n}\cdot\frac{\frac{n-1}{2}\,\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

$$= \sigma^2\,.\,\frac{n-1}{n}$$

$$\therefore \quad E\left\{\frac{ns^2}{n-1}\right\} = \sigma^2$$

$\therefore$ $\dfrac{ns^2}{n-1}$ is an unbiased estimate of σ^2.

17.2. Student's t-Distribution

Student's t statistic is defined by

$$t = \left(\frac{\bar{x}-\mu}{S}\right)\sqrt{n}$$

where $$S^2 = \frac{1}{n-1}\sum_{i-1}^{n}(x_i-\bar{x})^2 = \frac{ns^2}{n-1}$$

$$\therefore \quad \frac{t^2}{v} = \frac{(\bar{x}-\mu)^2}{s^2}, \text{ where } v = n-1$$

$$= \left\{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right\}^2 \Bigg/ \frac{ns^2}{\sigma^2}$$

Now, $\bar{x}$ is $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$

$\therefore$ $\left(\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2$ is a ψ^2 variate with 1 d.f.

and $\dfrac{ns^2}{\sigma^2}$ is a ψ^2 variate with $n-1 = v$ d.f.

$\therefore$ $\dfrac{t^2}{v}$ is a $\beta_2\left(\dfrac{v}{2}, \dfrac{1}{2}\right)$ variate

$\therefore$ Dist. of t is

$$dP = \frac{1}{\beta\left(\frac{v}{2},\frac{1}{2}\right)} \frac{\left(\frac{t^2}{v}\right)^{\frac{1}{2}-1}}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}} d\left(\frac{t^2}{v}\right), 0 < t^2 < \infty$$

$$= \frac{1}{\sqrt{v}\,.\,\beta\left(\frac{v}{2},\frac{1}{2}\right)} \frac{dt}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}} \quad -\infty < t < \infty$$

This distribution is known as student's t-distribution with v d.f.

Remark. t-distribution was first found by W.S. Gosset in 1908 in his paper entitled. '**The probable error of the mean**' written under the name of his student. Student defined his statistic as

$$t = \frac{\bar{x}-\mu}{S}$$

and investigated its sampling distribution. Later on in 1926, Prof. R.A. Fisher defined his own statistic and gave a rigorous proof for its sampling distribution. He defined his statistic as

$$t = \frac{\xi}{\sqrt{\frac{\psi^2}{n}}}$$

where ξ is a $N(0, 1)$, ψ^2 is a chi-square variate with n.d.f. and ξ, ψ^2 are independent.

Distribution of Fisher's t

Since ξ and ψ^2 are independ

$dP =$

Put $t =$

$\Rightarrow \quad \xi =$

$\therefore \quad \dfrac{\partial(\xi,\psi^2)}{\partial(t,u)} =$

$=$

$\therefore$ The joint distribution

$dP =$

$=$

$\therefore$ Marginal distribution

$\dfrac{1}{\sqrt{n}\,.\,\sqrt{2\pi}}\,.$

$= \dfrac{\quad}{\sqrt{n}\,\sqrt{2\pi}}$

where $y =$

$\frac{ns^2}{n-1}$

$n-1$

1

$d\left(\frac{t^2}{\nu}\right), 0 < t^2 < \infty$

$\frac{dt}{\left(\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \quad -\infty < t < \infty$

n with ν d.f.
osset in 1908 in his paper entitled.
ame of his student. Student defined

?6, Prof. R.A. Fisher defined his own
bution. He defined his statistic as

l.f. and ξ, ψ^2 are independent.

Distribution of Fisher's t

Since ξ and ψ^2 are independent, their joint distribution is

$$dP = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2} \cdot \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}\psi^2} (\psi^2)^{\frac{n}{2}-1}\, d\xi d\psi^2$$

$$-\infty < \xi < \infty,\ 0 \le \psi^2 < \infty$$

Put $\quad t = \dfrac{\xi}{\sqrt{\psi^2/n}}, \quad u = \psi^2$

$\Rightarrow \quad \xi = t\dfrac{\sqrt{u}}{\sqrt{n}}, \quad \psi^2 = u$

$$\therefore \quad \frac{\partial(\xi,\psi^2)}{\partial(t,u)} = \begin{vmatrix} \frac{\partial \xi}{\partial t} & \frac{\partial \xi}{\partial u} \\ \frac{\partial \psi^2}{\partial t} & \frac{\partial \psi^2}{\partial u} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{\sqrt{u}}{\sqrt{n}} & \frac{t}{2\sqrt{u}\sqrt{n}} \\ 0 & 1 \end{vmatrix} = \frac{\sqrt{u}}{\sqrt{n}}$$

$\therefore$ The joint distribution of t and u is

$$dP = \frac{1}{\sqrt{2\pi}\, 2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}u\left(1+\frac{t^2}{n}\right)} u^{\frac{n}{2}-1} \cdot \frac{\sqrt{u}}{\sqrt{n}}\, du\, dt$$

$$= \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{2\pi}\, 2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}u\left(1+\frac{t^2}{n}\right)} . u^{\frac{n}{2}-\frac{1}{2}}\, du\, dt$$

$$-\infty < t \infty,\ 0 < u < \infty$$

$\therefore$ Marginal distribution of t is

$$\frac{1}{\sqrt{n}.\sqrt{2\pi}} \cdot \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} (dt) \int_0^\infty e^{-\frac{1}{2}u\left(1+\frac{t^2}{n}\right)} u^{\frac{n}{2}-\frac{1}{2}}\, du$$

$$= \frac{dt}{\sqrt{n}\sqrt{2\pi}\, 2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-y} \left(\frac{2y}{1+\frac{t^2}{n}}\right)^{\frac{n-1}{2}} \left(\frac{2dy}{1+\frac{t^2}{n}}\right)$$

where $y = \dfrac{1}{2}u\left(1+\dfrac{t^2}{n}\right)$

$$= \frac{dt}{\sqrt{n}\,.\,\sqrt{2\pi}\,2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)}\;\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\,2^{\frac{n+1}{2}}\int_0^{\infty} e^{-y}.y^{\frac{n+1}{2}-1}\,dy$$

$$= \frac{dt}{\sqrt{n}\,.\sqrt{\pi}\;\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\,\Gamma\left(\frac{n+1}{2}\right)$$

$$= \frac{1}{\sqrt{n}}\;\frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right)}\;\frac{dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

which is t-distribution with n.d.f.

(2) Taking

$$\xi = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,\,1)$$

$$\psi^2 = \frac{ns^2}{\sigma^2}$$

which is chi-square variate with $(n-1)$ d.f. Fisher's statistic t takes the form

$$t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\;\frac{1}{\sqrt{\dfrac{ns^2}{\sigma^2\,(n-1)}}}$$

$$= \left(\frac{\bar{x}-\mu}{S}\right)\sqrt{n}$$

which is student's t statistic. Thus, student's t can be regarded as a particular case of Fisher's t.

17.2-1. Properties of t-distribution

$$t = \text{Mean} = E(t)$$

$$= \frac{1}{\sqrt{v}\,\beta\left(\frac{v}{2},\frac{1}{2}\right)}\int_{-\infty}^{\infty}\frac{t}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}\,dt = 0$$

$$\therefore \quad \mu_{2r+1} = E(t-\bar{t})^{2r+1}$$

$$= E(t^{2r+1})$$

$$= \frac{1}{\sqrt{v}\,\beta\left(\frac{v}{2},\frac{1}{2}\right)}\int_{-\infty}^{\infty}\frac{t^{2r+1}}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}\,dt = 0.$$

and $\quad \mu_{2r} = E(t^{2r})$.

=

=

This converges if $2r < v$. So if

Put $\quad \frac{t^2}{v} =$

$\therefore \quad 2t\,dt =$

$\therefore \quad \mu_{2r} =$

=

=

=

=

=

Put $\quad r =$

$\therefore \quad \mu_2 =$

$\therefore \quad \mu_4 =$

$\therefore \quad \beta_2 =$

$$= \frac{1}{\sqrt{\nu}\,\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_{-\infty}^{\infty} \frac{t^{2r}}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}\, dt$$

$$= \frac{2}{\sqrt{\nu}\,\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_{0}^{\infty} \frac{t^{2r}}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}\, dt$$

This converges if $2r < \nu$. So if $r < \frac{\nu}{2}$, μ_{2r} exist.

Put $\qquad \frac{t^2}{\nu} = x$

$\therefore \qquad 2t\, dt = \nu\, dx$

$$\therefore \qquad \mu_{2r} = \frac{\nu^{r+\frac{1}{2}}}{\sqrt{\nu}\,\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_{0}^{\infty} \frac{x^{r-\frac{1}{2}}}{(1+x)^{\frac{\nu+1}{2}}}\, dx$$

$$= \frac{\nu^r}{\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_{0}^{\infty} \frac{x^{r+\frac{1}{2}-1}}{(1+x)^{\left(\frac{\nu}{2}-r\right)+\left(r+\frac{1}{2}\right)}}\, dx$$

$$= \frac{\nu^r}{\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)}\, \beta\left(\frac{\nu}{2} - r, r + \frac{1}{2}\right)$$

$$= \frac{\nu^r}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}\, \Gamma\left(\frac{\nu}{2} - r\right)\Gamma\left(r + \frac{1}{2}\right)$$

$$= \frac{\left(r-\frac{1}{2}\right)\left(r-\frac{3}{2}\right)\cdots\frac{1}{2}}{\left(\frac{\nu}{2}-1\right)\left(\frac{\nu}{2}-2\right)\cdots\left(\frac{\nu}{2}-r\right)}\, \nu^r$$

$$= \frac{(2r-1)(2r-3)\cdots 1}{(\nu-2)(\nu-4)\cdots(\nu-2r)}\, \nu^r$$

Put $\qquad r = 1, 2$

$$\therefore \qquad \mu_2 = \frac{1}{\nu-2}\cdot \nu = \frac{\nu}{\nu-2} > 1$$

$$\therefore \qquad \mu_4 = \frac{3\cdot 1}{(\nu-2)(\nu-4)}\,\nu^2 = \frac{3\nu^2}{(\nu-2)(\nu-4)}$$

$$\therefore \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(\nu-2)}{\nu-4} \to 3 \text{ as } \nu \to \infty$$

$$\therefore \qquad \gamma_2 = \beta_2 - 3 = \frac{6}{v-4} \to 0 \text{ as } v \to \infty$$

Also,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0, \ \gamma_1 = \sqrt{\beta_1} = 0$$

Recurrence formula for moments

$$\mu_{2r} = \frac{(2r-1)\ldots\ldots(1)}{(v-2)\ldots\ldots(v-2r)} v^r$$

and

$$\mu_{2r-2} = \frac{(2r-3)\ldots\ldots 1}{(v-2)\ldots\ldots(v-2r+2)} v^{r-1}$$

Dividing

$$\frac{\mu_{2r}}{\mu_{2r-2}} = \frac{2r-1}{v-2r} \cdot v$$

17.2-2. Chief Features of the t-Probability Curve

The equation of the t-probability curve is

$$y = \frac{1}{\sqrt{v}\,\beta\left(\frac{v}{2}, \frac{1}{2}\right)} \frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}$$

(1) Since on changing t to $-t$, y does not change, curve is symmetrical about $t = 0$

Median $= 0$

(2) $y \to 0$ as $|t| \to \infty$.

$\therefore$ Curve is asymptotic to t-axis at both ends

(3) y decrease rapidly as $|t|$ increases.

$\therefore$ y is maximum for $t = 0$

$\therefore$ Mode $= 0$

Mean, Mode and Median coincide.

17.2-3. Limiting form of t-distribution

Density f^n of t - dist. is

$$f(t) = \frac{1}{\sqrt{v}\,\beta\left(\frac{v}{2}, \frac{1}{2}\right)} \frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}$$

$$= \frac{1}{\sqrt{v}} \cdot \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{v}{2}\right)} \frac{1}{\left\{\left(1+\frac{t^2}{v}\right)^{\frac{v}{t^2}}\right\}^{\frac{t^2}{2}}} \frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{1}{2}}} \qquad \ldots(1)$$

for large v,

$$\Gamma\left(\frac{v+1}{2}\right) = \left(\frac{v+1}{2} - 1\right)!$$

$= \left(\frac{v-1}{2}\right.$

$\simeq \sqrt{2\pi}\, e$

$\Gamma\left(\frac{v}{2}\right) =$

$\therefore \quad \frac{1}{\sqrt{v}} \cdot \frac{\left|\frac{v+1}{2}\right.}{\left|\frac{v}{2}\right.} \simeq$

$\to$

Also $\Gamma\left(\frac{1}{2}\right) =$

$\therefore \quad (1) \Rightarrow \lim f(t) = \frac{1}{\sqrt{2\pi}}$

which is the density f^n of a stand

$\therefore$ t-dist. becomes normal w

Example. *If x is t-distribute*

1

$r-1$

$\frac{+1}{2}$

ve is symmetrical about $t = 0$

$$\frac{1}{\;\;\;\frac{t^2}{2}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{1}{2}}} \quad ...(1)$$

$$= \left(\frac{\nu-1}{2}\right)$$

$$\simeq \sqrt{2\pi}\, e^{-\left(\frac{\nu-1}{2}\right)} \left(\frac{\nu-1}{2}\right)^{\frac{\nu-1}{2}+\frac{1}{2}}$$

$$= \sqrt{2\pi}\, e^{-\frac{\nu}{2}+\frac{1}{2}} \left(\frac{\nu-1}{2}\right)^{\frac{\nu}{2}}$$

$$\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2}-1\right)!$$

$$= \sqrt{2\pi}\, e^{-\frac{\nu}{2}+1} \cdot \left(\frac{\nu}{2}-1\right)^{\frac{\nu}{2}-1+\frac{1}{2}}$$

$$= \sqrt{2\pi}\, e^{-\frac{\nu}{2}+1} \cdot \left(\frac{\nu}{2}-1\right)^{\frac{\nu}{2}-\frac{1}{2}}$$

$$\therefore \quad \frac{1}{\sqrt{\nu}} \cdot \frac{\boxed{\frac{\nu+1}{2}}}{\boxed{\frac{\nu}{2}}} \simeq \frac{1}{\sqrt{\nu}} \cdot \frac{1}{\sqrt{e}} \cdot \frac{\left(\frac{\nu-1}{2}\right)^{\frac{\nu}{2}}}{\left(\frac{\nu}{2}-1\right)^{\frac{\nu}{2}-\frac{1}{2}}}$$

$$= \frac{1}{\sqrt{e}} \cdot \frac{\left(1-\frac{1}{\nu}\right)^{\frac{\nu}{2}}}{\left(1-\frac{2}{\nu}\right)^{\frac{\nu}{2}}} \cdot \left(\frac{1}{2}-\frac{1}{\nu}\right)^{\frac{1}{2}}$$

$$= \frac{1}{\sqrt{e}} \frac{\left\{\left(1-\frac{1}{\nu}\right)^{-\nu}\right\}^{-\frac{1}{2}}}{\left\{\left(1-\frac{2}{\nu}\right)^{-\frac{\nu}{2}}\right\}^{-1}} \cdot \left(\frac{1}{2}-\frac{1}{\nu}\right)^{\frac{1}{2}}$$

$$\to \quad \frac{1}{\sqrt{e}} \cdot \frac{e^{-1/2}}{e^{-1}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}$$

Also $\qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$\therefore \quad (1) \Rightarrow \lim f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}t^2}$$

which is the density f^n of a standard normal variate.

$\therefore$ t-dist. becomes normal when ν is large.

Example. *If x is t-distributed with k degrees of freedom, show that*

$$\frac{1}{1+\frac{x^2}{k}}$$

has a beta distribution.

17.3. t-tests

Tests of significance based on t-distribution are called t-tests. Various t-tests are :

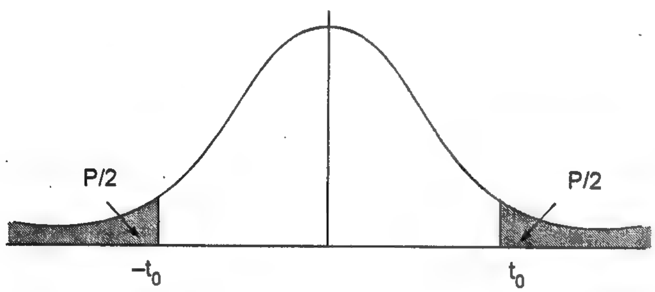
(*i*) Test for single proportion.
(*ii*) Test for the difference of means.
(*iii*) Test for the significance of an observed sample correlation co-efficient.
(*iv*) Test for the significance of an observed regression co-efficient.
(*v*) Test for the significance of a rank correlation co-efficient.

All these tests are for small samples and are based on fundamental assumption that the parent population is normal.

Rules of Decision

Let
$$P = P\{|t| > t_0\}$$
$$= 2P\ \{t > t_0\}$$

For various fixed values of P and for v ranging from 1 to 60; values of t have been tabulated in the form of t-tables. For $v > 60$, t is considered as a standard normal variate. The value t_0 is called the critical value of t at level of significance P and d.f. v.



To test the significance the calculated value of t is compared with tabulated value at certain specified level of significance. Generally 5% or 1% level are taken.

If calculated value of $|t|$ exceeds tabulated value, the null hypothesis is rejected and the difference is said to be significant and if it is less than tabulated value, the hypothesis is accepted at the level of significance adopted.

Remark

In above rules both the ends of t curve are considered and hence tests with these rules are called two-tailed tests. If, however, one tail is used tests are called single-tailed tests.

Since t-curve is symmetrical about $t = 0$
$$P\{t \geq t_0\,(\alpha)\} = P\{t \leq -\,t_0\,(\alpha)\}$$
where α is the level to significance and $t_0\,(\alpha)$ the critical value of t at level of significance α.
$$\therefore \quad \alpha = P\{|t| > t_0\,(\alpha)\} = 2P\{t \geq t_0\,(\alpha)\}$$
$$\Rightarrow \quad \frac{\alpha}{2} = P\{t \geq t_0\,(\alpha)\}$$
changing α to $2\,\alpha$
$$\Rightarrow \quad \alpha = P\{t \geq t_0\,(2\alpha)\}$$

Hence for a single tailed test, the critical values of t for level of significance α can be obtained from those of two tailed test by looking the values at level of significance 2α.

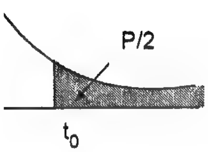
17.3.1. Test for Single Mean

Let $x_1, x_2, \ldots\, x_n$ be a random sample from a normal population with mean μ. The problem here is to test "is the sample mean differs significantly from the population mean μ" ? Assuming the null hypothesis. "There is no significant difference between the sample mean and the population mean", the statistic.

:d t-tests. Various t-tests are :

e correlation co-efficient.
sion co-efficient.
:o-efficient.
ı fundamental assumption that the

m 1 to 60; values of t have been
l as a standard normal variate. The
ance P and d.f.v.



compared with tabulated value at
.% level are taken.
null hypothesis is rejected and the
tabulated value, the hypothesis is

:d and hence tests with these rules
:sts are called single-tailed tests.

ralue of t at level of significance α.

for level of significance α can be
ues at level of significance 2α.

nal population with mean μ. The
ficantly from the population mean
ant difference between the sample

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

where $\bar{x}$ = sample mean

and $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{n}{n-1}s^2 \quad \text{(where } s^2 \text{ is sample s.d.)}$$

which follows student's t-distribution with $(n-1)$ d.f., is calculated.

Ex. 17-1. *A mechanist is making engine parts with axle diameters 0·700″. A random sample of 10 parts shows a mean diameter of 0·742″ with a s.d. of 0·04″. Test whether work is meeting the specification.*

Sol. Here $\bar{x} = 0{\cdot}742$, $n = 10$, $s = 0{\cdot}04$ and $\mu = 0{\cdot}7''$.

$$\therefore \quad t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \left(\frac{\bar{x}-\mu}{s}\right)\sqrt{n-1} \simeq 3{\cdot}15.$$

No. of $d.f. = 10 - 1 = 9$

From tables $t_{0\cdot05}$ for 9 $d.f. = 2{\cdot}26$

$\therefore \quad t_{\text{cal}} > t_{0\cdot05}$

$\therefore$ $\bar{x}$ differs from μ significantly and hence the work is not meeting the specification.

Ex. 17-2. *Ten individuals are chosen at random from a normal population and the heights are found to be inches 63, 63, 66, 67, 68, 69, 70, 70, 71 and 71. In the light of these data discuss the suggestion that the mean height in the universe is 66″ having given that*

$$\left.\begin{array}{ll} \textit{for } t = 1{\cdot}8 & P = 0{\cdot}947 \\ \textit{for } t = 1{\cdot}9 & P = 0{\cdot}955 \end{array}\right\} \text{ for 9 } d.f.$$

where P is the area to the left of the ordinate at t.

Sol. Calculation of mean and s.d.

											Total
x :	63	63	66	67	68	69	70	70	71	71	
$X = x - 68$:	–5	–5	–2	–1	0	1	2	2	3	3	–2
X^2 :	25	25	4	1	0	1	4	4	9	9	82

$$\therefore \quad \bar{x} = 68 + \left(-\frac{2}{10}\right) = 67{\cdot}8,\ s^2 = \frac{82}{10} - \left(-\frac{2}{10}\right)^2 = 8{\cdot}2 - 0{\cdot}04 = 8{\cdot}16$$

Assuming population mean to be 66, $\mu = 66$

$$t = \left(\frac{\bar{x}-\mu}{s}\right)\sqrt{n-1} = \frac{(1{\cdot}8)\sqrt{9}}{\sqrt{8{\cdot}16}} = \frac{5{\cdot}4}{\sqrt{8{\cdot}16}} \cong 1{\cdot}89$$

when t increases by 0·1, P increases by 0·008

when t increases by 0·09 P increases by $\left(\frac{0{\cdot}008}{0{\cdot}1}\right)(0{\cdot}09) = 0{\cdot}0072$

$\therefore$ For $t = 1{\cdot}89$, $\quad P = 0{\cdot}9542$

$\therefore$ $P_F = 2(1-P) = 2(1-0{\cdot}9542) = 0{\cdot}0916 > 0{\cdot}05$

$\therefore$ Difference is not significant at 5% level of significance and hence test provides no evidence against the population mean being 66″

Ex. 17-3. *Nine patients, to whom a certain drug was administrated, registered the following increments in blood pressure :*

7, 3, – 1, 4, – 3, 5, 6, – 4, 1

Show that the data do not indicate that the drug was responsible for these increments. The values of t for 10, 9 and 8 d.f. at 5% level of significance are 2·23, 2·26 and 2·31 respectively.

Sol. Let x be the variable for the increment in blood pressure.

										Total
x :	7	3	− 1	4	− 3	5	6	− 4	1	18
$x-\bar{x}$:	5	1	− 3	2	− 5	3	4	− 6	− 1	
$(x-\bar{x})^2$:	25	1	9	4	25	9	16	36	1	126

$$\bar{x} = \frac{18}{9} = 2, \quad S^2 = \frac{1}{9-1}(126) \approx 15{\cdot}75.$$

Assuming that the drug was not responsible for the increments in blood pressure, $\mu = 0$

$$t = \frac{2\sqrt{9}}{\sqrt{15\cdot 75}} \simeq 1{\cdot}51$$

No. of d.f. = 9 − 1 = 8

$\therefore$ $t_{0\cdot 05} = 2{\cdot}31$

$\therefore$ $t_{cal} < t_{0\cdot 05}$.

$\therefore$ The data do not indicate that the drug was responsible for increment in blood pressure.

Ex. 17-4. *Ten patients to whom a drug administered registered the following additional hours of sleep :*

0·7, − 1·1, − 0·2, 1·2, 0·1, 3·4, 3·7, 0·8, 1·8, 2·0

Compute the statistic you would use to determine whether these data justify the claim that the drug does produce additional sleep.

Sol. **Calculation of mean and s.d.**

x :	0·7	− 1·1	− 0·2	1·2	0·1	3·4	3·7	0·8	1·8	2·0	12·4
$x-\bar{x}$:	− 0·54	− 2·34	− 1·44	− 0·04	− 1·14	2·16	2·46	− 0·44	0·56	0·76	
$(x-\bar{x})^2$:	0·2916	5·4756	2·0736	0·0016	1·2996	4·6656	6·0516	0·1936	0·3136	0·5776	20·9440

$$\bar{x} = \frac{12\cdot 4}{10} = 1{\cdot}24, \; s^2 = 2{\cdot}0944$$

Assuming that drug does not produce additional sleep, $\mu = 0$

$$\therefore \qquad t = \frac{(1\cdot 24)\sqrt{9}}{\sqrt{2\cdot 0944}} \approx 2{\cdot}6.$$

Ex. 17-5. *A certain stimulus administered to each of the 12 patients resulted in the following increases of blood pressure :*

5, 2, 8, − 1, 3, 0, 6, − 2, 1, 5, 0, 4

Can it be concluded that stimulus will be, in general accompanied by an increase in blood pressure ? Given that $t_{0.05}(10) = 2{\cdot}23$, $t_{0\cdot 05}(11) = 2{\cdot}20$, $t_{0\cdot 05}(12) = 2{\cdot}18$ where $t_\alpha(n)$ denotes the value of 't' for 'n' d.f. at 'α' level of significance.

Sol. **Calculation of mean and s.d.**

													Total
x :	5	2	8	−1	3	0	6	−2	1	5	0	4	31
x^2 :	25	4	64	1	9	0	36	4	1	25	0	16	185

$\bar{x}$ =

Assuming that stimulus will ı
$\mu = 0$.

$\therefore$ t =

No. of $d.f.$ =

Now $t_{0\cdot 05}(11)$ =

$\therefore$ t_{cal} >

$\therefore$ Assumption is wrong and
an increase in blood pressure.

Ex. 17-6. *Show that the 95%*

$\bar{x} \pm \frac{St_{0\cdot 05}}{\sqrt{n}}$. *Deduce that for a rand*
the squares of the deviations fro
population, 95% fiducial limits for

Sol. Since $P\{|t| \le t_{0\cdot 05}\} = 0{\cdot}9$

$|t|$ =

i.e., $|\bar{x}-\mu| \le$

i.e., $\bar{x} - \frac{St_{0\cdot 05}}{\sqrt{n}} \le$

Here n = 1

No. of d.f. = 16 − 1 = 15

$\therefore$ $t_{0\cdot 05} = 2{\cdot}13$

$\therefore$ 95% fiducial limits are

$41{\cdot}5 \mp \frac{3}{4}(2{\cdot}13)$ *i.e.,* 3

1. A machine which produces m
turn out washers having a thi
washers has an average thic
significance of such a deviati
2. Find the "student's t" for the
- 4,
taking the mean of the univer
3. Find student's t for the follow
− 6,
taking μ to be zero.
4. Ten individuals are chosen at
63, 63, 66, 67, 68, 69, 70, 70,

$$\bar{x} = \frac{31}{12} \quad \text{and} \quad s^2 = \frac{185}{12} - \left(\frac{31}{12}\right)^2 = \frac{1259}{144}$$

Assuming that stimulus will not be accompanied by an increase in blood pressure, $\mu = 0$.

$$\therefore \qquad t = \frac{31}{12} \times \frac{\sqrt{11}}{\sqrt{\frac{1259}{144}}} = \frac{31\sqrt{11}}{\sqrt{1259}} \cong 2{\cdot}9$$

$$\text{No. of } d.f. = 12 - 1 = 11$$

Now $\qquad t_{0{\cdot}05}(11) = 2{\cdot}20$

$\therefore \qquad t_{\text{cal}} > t_{0{\cdot}05}$

$\therefore$ Assumption is wrong and hence the stimulus will be, in general, accompanied by an increase in blood pressure.

Ex. 17-6. *Show that the 95% fiducial limits for the mean* μ *of the population are* $\bar{x} \pm \dfrac{St_{0{\cdot}05}}{\sqrt{n}}$. *Deduce that for a random sample of* 16 *values with mean* 41·5″ *and the sum of the squares of the deviations from the mean* 135 *(inches)*^{2} *and drawn from a normal population, 95% fiducial limits for the mean of the population are* 39·9″ and 43·1″.

Sol. Since $P\{|t| \le t_{0{\cdot}05}\} = 0{\cdot}95$, 95%, fiducial limits for the mean μ are given by

$$|t| = \left|\frac{\bar{x}-\mu}{S}\sqrt{n}\right| \le t_{0{\cdot}05}$$

i.e.,
$$|\bar{x} - \mu| \le \frac{St_{0{\cdot}05}}{\sqrt{n}}$$

i.e.,
$$\bar{x} - \frac{St_{0{\cdot}05}}{\sqrt{n}} \le \mu \le \bar{x} + \frac{St_{0{\cdot}05}}{\sqrt{n}}$$

Here $\qquad n = 16, \ \bar{x} = 41{\cdot}5, \ S^2 = \dfrac{135}{16-1} = 9$

No. of d.f. = 16 − 1 = 15

$\therefore \quad t_{0{\cdot}05} = 2{\cdot}13$

$\therefore \quad$ 95% fiducial limits are

$$41{\cdot}5 \mp \frac{3}{4}(2{\cdot}13) \ \textit{i.e.,} \ 39{\cdot}9 \text{ and } 43{\cdot}1.$$

EXERCISES

1. A machine which produces mica insulating washers of use in electric devices is set to turn out washers having a thickness of 10 mils (1 mil = 0·001 inch). A sample of 10 washers has an average thickness of 9·52 mils with a s.d. of 0·60 mil. Test the significance of such a deviation. **[Ans.** $t = 2{\cdot}4$**]**
2. Find the "student's t" for the following variate values in a sample of eight

   − 4, − 2 − 2, 0, 2, 2, 3, 3

   taking the mean of the universe to be zero. **[Ans.** 0·3**]**
3. Find student's t for the following variate values in a sample of ten :

   − 6, − 4, − 3, − 2, − 2, 0, 1, 1, 3, 5

   taking μ to be zero. **[Ans.** 0·7**]**
4. Ten individuals are chosen at random from a population, their heights are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71 and 71 inches respectively. Test whether the mean

height is 69·6″ in the population, given that for 9 d.f. $P\{|t| \geq 2 \cdot 262\} = 0 \cdot 05$.
[Ans. $|t| = 1 \cdot 89$**]**

5. The nine items of a sample had the following values :

45, 47, 50, 52, 48, 47, 49, 53, 51

Does the mean of the nine items differ significantly from the assumed population mean of 47·5 ? Given that

$$v = 8 \begin{cases} P = 0 \cdot 945 \text{ for } t = 1 \cdot 8 \\ P = 0 \cdot 953 \text{ for } t = 1 \cdot 9 \end{cases}$$
[Ans. $t = 1 \cdot 84$**]**

6. The gain (in bushels per acre) in yield due to the use of a variety of wheat in nine plots is as follows :

16·3, 13·4, 3·8, 7·9, 2·6, 2·5, 9·6, 7·2 and 3·3

Are the observations consistent with the hypothesis that the average gain is 7·5 bushels per acre ? **[Ans. Yes]**

7. Ten individuals are chosen at random from a population and their heights are found to be inches 63, 63, 64, 65, 66, 69, 69, 70, 70 and 71. Discuss the suggestion that the mean height in the universe is 65 inches. (Given that for 9 *d.f.* $t_{0 \cdot 05} = 2 \cdot 262$)
[Ans. $t = 2 \cdot 02$**]**

8. The table signifies additional hours of sleep gained by 10 patients in an experiment with a sleeping drug :

Patient	1	2	3	4	5	6	7	8	9	10
Hours gained	0·7	− 1·1	− 0·2	− 1·2	0·1	3·4	3·7	0·8	1·9	2·0

Assuming that the hours of sleep is a normally distributed variable, calculate 't' for the above table. **[Ans. 1·9]**

9. A certain drug caused the following increases in blood pressure of 12 patients.

3, 0, 6, − 2, 1, 5, 2, 8, 0, − 1, 1, 5

Can it be concluded that the stimulus does not effect blood pressure ? **[Ans.** $t = 2 \cdot 6$**]**

10. The mean weekly sale of the ice cream bar was 146·3 bars. After an advertising campaign the mean weekly sale in 22 shops for a typical week increased to 153·7 and showed a standard deviation 17·2. Is this evidence that the advertising was successful ? (Given that for d.f. = 21, $t_{0 \cdot 05} = 2 \cdot 08$] **[Ans.** $t = 1 \cdot 97$**]**

11. A certain colliery is supposed to supply coal of ash content about 15. To test this 20 random samples of the colliery's coal are selected and tested. The null hypothesis is that the ash content is in fact 15. The results of 20 tests gave an average ash content of 16·8 with a standard deviation of 3·6. Is this sufficient evidence for rejecting the hypothesis ?
(Given that for 19 d.f. $t_{0 \cdot 05} = 2 \cdot 09$) **[Ans.** $t = 2 \cdot 18$**]**

12. A sample of 11 rats from a central population had an average blood viscosity of 3·92 with a *s.d.* of 0·61. On the basis of this sample establish 95% confidence limits of μ, the mean blood viscosity of central population. **[Ans. 3·51 and 4·33]**

13. The average breaking strength of steel rods is specified to be 17·5 lbs. To test this a sample of 14 rods was tested and gave the following results (in unit of 1,000 lbs) :

15, 18, 16, 21, 19, 21, 17, 17, 15, 17, 20, 19, 17, 18

Is the result of the experiment significant? Also obtain the 95% fiducial limits from the sample for the average breaking strength of steel rods. **[Ans.** $t = 0 \cdot 68$**]**

14. The mean of I.Q's of 10 boys is 97·2 with the sum of the squares of the deviation from the mean of 1833·6. Do these data support the assumption of a population mean I.Q's of 100 ? Find the 95% confidence limits for the population mean.
[Ans. $t = 0 \cdot 62$, **107·41 and 86·99]**

d.f. $P\{|t| \geq 2\cdot262\} = 0\cdot05$.

[**Ans.** $|t| = 1\cdot89$]

ues :

19, 53, 51

intly from the assumed population

$=1\cdot8$

[**Ans.** $t = 1\cdot84$]

$=1\cdot9$

se of a variety of wheat in nine plots

and $3\cdot3$

s that the average gain is $7\cdot5$ bushels

[**Ans. Yes**]

ilation and their heights are found to

71. Discuss the suggestion that the

that for 9 *d.f.* $t_{0\cdot05} = 2\cdot262$)

[**Ans.** $t = 2\cdot02$]

ied by 10 patients in an experiment

6	7	8	9	10
3·4	3·7	0·8	1·9	2·0

istributed variable, calculate 't' for

[**Ans.** 1·9]

blood pressure of 12 patients.

1, 1, 5

ect blood pressure ? [**Ans.** $t = 2\cdot6$]

is 146·3 bars. After an advertising

typical week increased to 153·7 and

that the advertising was successful ?

[**Ans.** $t = 1\cdot97$]

ish content about 15. To test this 20

d and tested. The null hypothesis is

tests gave an average ash content of

ufficient evidence for rejecting the

[**Ans.** $t = 2\cdot18$]

d an average blood viscosity of 3·92

stablish 95% confidence limits of μ,

[**Ans.** 3·51 and 4·33]

ecified to be 17·5 lbs. To test this a

ving results (in unit of 1,000 lbs) :

), 17, 18

obtain the 95% fiducial limits from

steel rods. [**Ans.** $t = 0\cdot68$]

of the squares of the deviation from

sumption of a population mean I.Q's

population mean.

Ans. $t = 0\cdot62$, 107·41 and 86·99]

15. A sample of 20 items has mean 42 units and standard deviation 5 units. Test the assumption that it is a random sample from a normal population with mean 45 units. Also obtain 95% fiducial limits.
(Given that $t_{0\cdot05} = 2\cdot09$ for 19 d.f.) [**Ans.** $t = 2\cdot6$, 39·6 and 44·4]

16. A drug was administered to 10 patients and the increments in their B.P. were recorded to be

0, 3, – 2, 4, – 3, 4, 6, 0, 0, 2

Is it reasonable to believe that the drug has no effect on B.P. ?

17. The weights of 15 bags of sugar taken from the filling machine are given below (in kgs)

16·1, 15·8, 15·9, 16·1, 16·2, 16·0, 15·9, 16·0, 15·7, 15·7, 15·8, 16·0, 16·0, 15·8, 15·7

If the machine is supposed to be giving 16 k.g. of sugar per bag on an average, test whether the data suggest that the machine is failing in its purpose.

17.3-2. Test for the difference of means

Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2, \ldots y_{n_2}$ be two independent random samples with means $\bar{x}$, $\bar{y}$ and standard deviations s_1, s_2 respectively from two normal populations with the *same variance* σ. The problem is to test the hypothesis that the population means are μ_1, and μ_2. Assuming the population means μ_1 and μ_2, the statistic is defined as

$$t = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

where

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2\right]$$

$$= \frac{1}{n_1 + n_2 - 2}\left[n_1 s_1^2 + n_2 s_2^2\right]$$

Now $\qquad \bar{x} \sim N(\mu_1, \sigma/\sqrt{n_1})$

and $\qquad \bar{y} \sim N(\mu_2, \sigma/\sqrt{n_2})$

$$\therefore \quad \bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\therefore \quad \left\{\frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}\right\}^2 \text{ is a } \psi^2 - \text{variate with 1 } d.f.$$

Also $n_1 \dfrac{s_1^2}{\sigma^2}$ is a ψ^2 – variate with $(n_1 - 1)$ *d.f.*

and $\quad n_2 \dfrac{s_2^2}{\sigma^2}$ is a ψ^2 – variate with $(n_2 - 1)$ *d.f.*

$$\therefore \quad \frac{n_1 s_1^2}{\sigma^2} + \frac{n_2 s_2^2}{\sigma^2} = \frac{n_1 s_1^2 + n_2 s_2^2}{\sigma^2} = \frac{\nu S^2}{\sigma^2},$$

where $v = n_1 + n_2 - 2$, is a ψ^2 variate with $v = (n_1 + n_2 - 2)$ *d.f.*

$$\therefore \quad \left\{\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\right\}^2 \Bigg/ \frac{vS^2}{\sigma^2} = \frac{t^2}{v}$$

is $\qquad \beta_2\left(\dfrac{v}{2}, \dfrac{1}{2}\right)$ variate.

$\therefore$ Statistic t follows t-distribution with $v = n_1 + n_2 - 2$ *d.f.*

If the hypothesis to be tested is "Are the two population means same or the two sample means differ significantly", under the null hypothesis "population means are same *i.e.*, $\mu_1 = \mu_2$ or the two sample means do not differ significantly" the statistic to be calculated becomes

$$t = \frac{\bar{x}-\bar{y}}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

which also follows t-distribution with $(n_1 + n_2 - 2)$ *d.f.*

(*ii*) If (*i*) $n_1 = n_2 = n$(say) and (*ii*) the samples are not independent but the sample observations are paired together *i.e.*, the pair of observations (x_i, y_i) $(i = 1, 2, \ldots n)$ correspond to the same (*i*th) sample unit. The problem here again is to test "Are the sample means differ significantly".

Under the null hypothesis "sample means do not differ significantly" the statistic

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

where

$$d_i = x_i - y_i$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \text{ and}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

which follows t–distribution with $(n - 1)$ *d.f., is calculated.*

Ex. 17-7. *For a random sample of* 10 *pigs fed on diet A the increases in weight in pounds in a certain period were*

10, 6, 16, 17, 13, 12, 8, 14, 15,. 9 *lbs.*

For another sample of 12 *pigs, fed on diet B the increases in the same period were*

7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17 *lbs.*

Test whether diets A and B differ significantly as regard the effect on increases in weight (or test whether the mean increases in the two samples are significantly different). You may use the fact 5% *value of t* for 20 *d.f. is* 2·09.

Sol. **Calculation of mean and s.d.**

x :		10	6	16	17	13	12	8		14	15	9	Total
$X = x - 13$:		−3	−7	3	4	0	−1	−5		1	2	−4	−10
X^2 :		9	49	9	16	0	1	25		1	4	16	130
y :	7	13	22	15	12	14	18	8	21	23	10	17	
$Y = y - 14$:	−7	−1	8	1	−2	0	4	−6	7	9	−4	3	12
Y^2 :	49	1	64	1	4	0	16	36	49	81	16	9	326

2) *d.f.*

$n_2 - 2$ *d.f.*
tion means same or the two sample
opulation means are same *i.e.*, μ_1 =
ıe statistic to be calculated becomes

re not independent but the sample
ons (x_i, y_i) $(i = 1, 2, \dots n)$ correspond
o test "Are the sample means differ

iffer significantly" the statistic

ted.
n diet A the increases in weight in

5,. 9 lbs.
creases in the same period were
3, 10, 17 lbs.
ard the effect on increases in weight
ıre significantly different). You may

nd s.d.

8		14	15	9	Total
−5		1	2	−4	−10
25		1	4	16	130
8	21	23	10	17	
−6	7	9	−4	3	12
36	49	81	16	9	326

$$\therefore \quad \bar{x} = 13 + \frac{(-10)}{10} = 12, \quad \text{and} \quad \bar{y} = 14 + \left(\frac{12}{12}\right) = 15$$

$$\therefore \quad s_1^2 = \mu_2 \text{ for } x = \mu_2 \text{ for } X = \frac{\Sigma X^2}{10} - \left(\frac{\Sigma X}{10}\right)^2 = 13 - 1 = 12$$

and $$s_2^2 = \mu_2 \text{ for } y = \mu_2 \text{ for } Y = \frac{326}{12} - 1 = \frac{314}{12}$$

$$\therefore \quad S^2 = \frac{1}{10+12-2}\left\{(10)(12) + (12)\frac{(314)}{12}\right\} = 21 \cdot 7$$

Assume the diets *A* and *B* do not differ significantly as regard the effect, on increases in weight *i.e.*, the mean increases in two samples are not significantly different.

$$\text{Now} \qquad |t| = \frac{12 \sim 15}{\sqrt{21 \cdot 7}\sqrt{\frac{1}{10} + \frac{1}{12}}} = \frac{3\sqrt{120}}{\sqrt{21 \cdot 7}\sqrt{22}}$$

$$\approx 1 \cdot 5$$

$$\text{No. of d.f.} = 10 + 12 - 2 = 20$$

$$\text{Now} \qquad t_{0 \cdot 05} \text{ for } 20 \; d.f. = 2 \cdot 09$$

$$\therefore \qquad t_{\text{cal}} < t_{0 \cdot 05}$$

Assumption may be correct.

Ex. 17·8. *The following data represent the yield in bushels of Indian corn on ten sub-divisions of equal areas of two agricultural plots, in which Plot I was a central plot treated the same as Plot II except for the amount of phosphorus applied as a fertilizer* :

Plot I :	6·2	5·7	6·5	6·0	6·3	5·8	5·7	6·0	6·0	5·8
Plot II :	5·6	5·9	5·6	5·7	5·8	5·7	6·0	5·5	5·7	5·5

Is there a significant difference between the yields on the two plots, using the difference between their means as a criterion of judgment ?

											Total
x :	6·2	5·7	6·5	6·0	6·3	5·8	5·7	6·0	6·0	5·8	60
$X = x - \bar{x}$:	0·2	− 0·3	0·5	0	0·3	− 0·2	− 0·3	0	0	− 0·2	
X^2 :	0·04	0·09	0·25	0	0·09	0·04	0·09	0	0	0·04	0·64
y :	5·6	5·9	5·6	5·7	5·8	5·7	6·0	5·5	5·7	5·5	57
$Y = y - \bar{y}$:	− 0·1	0·2	− 0·1	0	0·1	0	0·3	− 0·2	0	− 0·2	
Y^2 :	0·01	0·04	0·01	0	0·01	0	0·09	0·04	0	0·04	0·24

$$\bar{x} = \frac{\Sigma x}{10} = 6 \quad \text{and} \quad \bar{y} = \frac{57}{10} = 5 \cdot 7$$

$$\therefore \quad S^2 = \frac{1}{10+10-2}(0 \cdot 64 + 0 \cdot 24) = \frac{0 \cdot 44}{9}$$

$$\therefore \quad t = \frac{0 \cdot 3}{\sqrt{\frac{0 \cdot 44}{9}}\sqrt{\frac{1}{10} + \frac{1}{10}}} \simeq 3 \cdot 03$$

$$\text{No. of } d.f. = 10 + 10 - 2 = 18$$

Now $\quad t_{0\cdot05}$ for 18 d.f. $= 2\cdot10$

$\therefore \quad t_{cal} > t_{0\cdot05}$

$\therefore$ The difference between the yields of the two plots is significant.

Ex. 17-9. *Two independent samples of* 8 *and* 7 *items respectively had the following values :*

Sample I :	9	11	13	11	15	9	12	14
Sample II :	10	12	10	14	9	8	10	

Is the difference between the means of the samples significant ?

Given that if $P\{|t| > t_0\} = 0\cdot05$, $t_0 = 2\cdot16$ *for* 13 *d.f. and* $t_0 = 2\cdot13$ *for* 15 *d.f.*

Sol. **Calculation of mean and s.d.**

									Total
x :	9	11	13	11	15	9	12	14	
$X = x - 11$:	-2	0	2	0	4	-2	1	3	6
X^2 :	4	0	4	0	16	4	1	9	38
y :	10	12	10	14	9	8	10		
$Y = y - 10$:	0	2	0	4	-1	-2	0		3
Y^2 :	0	4	0	16	1	4	0		25

$$\therefore \quad \bar{x} = 11 + \frac{6}{8} = \frac{47}{4} \quad \text{and} \quad \bar{y} = 10 + \frac{3}{7} = \frac{73}{7}$$

$$s_1^2 = \frac{38}{8} - \left(\frac{6}{8}\right)^2 = \frac{67}{16}$$

and
$$s_2^2 = \frac{25}{7} - \left(\frac{3}{7}\right)^2 = \frac{166}{49}$$

$$\therefore \quad S^2 = \frac{1}{8+7-2}\left\{8\left(\frac{67}{16}\right) + 7\left(\frac{166}{49}\right)\right\} = \frac{801}{(14)(13)}$$

$$\therefore \quad |t| = \frac{\frac{47}{4} \sim \frac{73}{7}}{\sqrt{\frac{801}{14\cdot13}}\sqrt{\frac{1}{8}+\frac{1}{7}}} = \frac{37}{\sqrt{(801)(15)}}\sqrt{13} \approx 1\cdot2$$

No. of *d.f.* $= 8 + 7 - 2 = 13$

$\therefore \quad t_{0\cdot05}(13) = 2\cdot16$

$\therefore \quad t_{cal} < t_{0\cdot05}$

$\therefore$ Difference is not significant.

Ex. 17-10. *Two horses A and B were tested according to the time (in seconds) to run a particular track with the following result :*

Horse A :	28	30	32	33	33	29	34
Horse B :	29	30	30	24	27	29	

Test whether you can discriminate between the two horses. You can use the fact that 5% *value of t for* 11 *d.f. is* 2·20.

Sol. Let x and y be the time variable for horses A and B respectively.

Calculation of mean and s.d.

								Total
x :	28	30	32	33	33	29	34	
$X = x - 32$:	-4	-2	0	1	1	-3	2	-5
X^2 :	16	4	0	1	1	9	4	35
y :	29	30	30	24	27	29		
$Y = y - 29$:	0	1	1	-5	-2	0		-5
Y^2 :	0	1	1	25	4	0		31

$$\therefore \quad \bar{x} = 32 - \frac{5}{7} = \frac{219}{7}, \quad \bar{y} = 29 - \frac{5}{6} = \frac{169}{6}$$

$$s_1^2 = \mu_2 \text{ of } x = \frac{35}{7} - \left(\frac{-5}{7}\right)^2 = 5 - \frac{25}{49} = \frac{220}{49}$$

$$s_2^2 = \mu_2 \text{ of } y = \frac{31}{6} - \left(\frac{-5}{6}\right)^2 = \frac{161}{36}$$

$$\therefore \quad S^2 = \frac{1}{7+6-2}\left\{\frac{220}{7} + \frac{161}{6}\right\} = \frac{2447}{462}$$

$$\therefore \quad |t| = \frac{\frac{219}{7} \sim \frac{169}{6}}{\sqrt{\frac{2447}{462}}\sqrt{\frac{1}{7} + \frac{1}{6}}} = \frac{131\sqrt{11}}{\sqrt{(13)(2447)}} = 2 \cdot 4$$

$$\text{No. of } d.f. = 7 + 6 - 2 = 11$$

$$\therefore \quad t_{0 \cdot 05} = 2 \cdot 20$$

$$\therefore \quad t_{cal} > t_{0 \cdot 05}$$

$\therefore$ Difference is significant and hence two horses can be discriminated.

Ex. 17-11. *The following table shows the mean number of bacterial colonies per plate obtainable by four slightly different methods from soil samples taken at 4 P.M. and 8 P.M. respectively.*

Methods	*A*	*B*	*C*	*D*
4 *P.M.*	29·75	27·50	30·25	27·80
8 *P.M.*	39·20	40·60	36·20	42·40

Are there significantly more bacteria at 8 P.M. than at 4 P.M.?

Sol. **Calculation of mean and s.d. of the difference**

Method	4 P.M. (x)	8 P.M. (y)	$d = y - x$	$d - \bar{d}$	$(d - \bar{d})^2$
A	29·75	39·20	9·45	$-1 \cdot 325$	1·756
B	27·50	40·60	13·10	2·325	5·406
C	30·25	36·20	5·95	$-4 \cdot 825$	23·281
D	27·80	42·40	14·60	3·825	14·631
Total			43·10		45·074

$$\bar{d} = \frac{43 \cdot 10}{4} = 10 \cdot 775, \quad S^2 = \frac{45 \cdot 074}{3}$$

$$\therefore \quad t = \frac{(10 \cdot 775)}{\sqrt{45 \cdot 074}} \sqrt{3(4)} = \frac{21 \cdot 55\sqrt{3}}{\sqrt{45 \cdot 074}}$$

$$= 5 \cdot 56$$

$$\text{No. of } d.f. = 4 - 1 = 3$$

$$\text{Now} \quad t_{0 \cdot 05}(3) = 3 \cdot 18 < t_{\text{cal}}$$

$\therefore$ Difference is highly significant and hence there are significantly more bacteria at 8 P.M. than at 4 P.M.

Ex. 17-12. *From the data given below test whether there is a significant difference between the effects of two drugs, on the assumption that different random samples of patients were used to test the two drugs A and B.*

Additional hours of sleep gained by use of soporific drugs.

Patient :	1	2	3	4	5	6	7	8	9	10
Drug A :	0·7	−1·6	−0·2	−1·2	−0·1	3·4	3·7	0·8	0	2·0
Drug B :	1·9	0·8	1·1	0·1	−0·1	4·4	5·5	1·6	4·6	3·6

Sol. **Calculation of mean and s.d. of the difference**

Patient	x	y	$d = y - x$	$(d - \bar{d})$	$(d - \bar{d})^2$
1	0·7	1·9	1·2	−0·4	0·16
2	−1·6	0·8	2·4	0·8	0·64
3	−0·2	1·1	1·3	−0·3	0·09
4	−1·2	0·1	1·3	−0·3	0·09
5	−0·1	−0·1	0	−1·6	2·56
6	3·4	4·4	1·0	−0·6	0·36
7	3·7	5·5	1·8	0·2	0·04
8	0·8	1·6	0·8	−0·8	0·64
9	0	4·6	4·6	3·0	9·00
10	2·0	3·6	1·6	0	0·00
Total			16·0		13·58

$$\bar{d} = \frac{16}{10} = 1 \cdot 6, \quad S^2 = \frac{13 \cdot 58}{9}$$

$$\therefore \quad t = \frac{1 \cdot 6}{\sqrt{\frac{13 \cdot 58}{9}}} \cdot \sqrt{10} = \frac{4 \cdot 8\sqrt{10}}{\sqrt{13 \cdot 58}} = 4.12$$

$$\text{No. of } d.f. = 10 - 1 = 9$$

$$\therefore \quad t_{0 \cdot 05}(9) = 2 \cdot 26 < t_{\text{cal}}$$

$\therefore$ Difference is highly significant.

Ex. 17-13. *In a certain experiment to compare two types of pig-foods A and B, the following result of increase in weights were observed in pigs :*

Pig No. :
Increase in *Food A* :
w.t. in lbs. *Food B* :

(*a*) *Assuming that two sam*
better than food A ? (b) Examin
the foods.

Sol. (*a*) In this case the tw
of means test for unpaired data
Let x and y be the increase

Food A	
x	$X = x - 50$
49	−1
53	3
51	1
52	2
47	−3
50	0
52	2
53	3
Total	7

$\therefore \quad \bar{x} =$

$\Sigma(x - \bar{x})^2 =$

$=$

$=$

$\therefore \quad S^2 =$

$\therefore \quad t =$

No. of $d.f.$ =

$\therefore \quad t_{0 \cdot 05}$ =

$\therefore$ At 5% level, difference
is better.

(*b*) In this case the two sampl
means test for paired data will be

Pig No.		:	1	2	3	4	5	6	7	8
Increase in	*Food A*	:	49	53	51	52	47	50	52	53
w.t. in lbs.	*Food B*	:	52	55	52	53	50	54	54	53

(*a*) *Assuming that two samples of pigs are independent can we conclude that food B is better than food A ?* (*b*) *Examine the case when the same set of eight pigs were used in both the foods.*

Sol. (*a*) In this case the two samples of pigs are independent and hence the difference of means test for unpaired data can be applied.

Let x and y be the increase in weights due to foods A and B respectively.

Food A			*Food B*		
x	$X = x - 50$	X^2	y	$Y = y - 52$	Y^2
49	-1	1	52	0	0
53	3	9	55	3	9
51	1	1	52	0	0
52	2	4	53	1	1
47	-3	9	50	-2	4
50	0	0	54	2	4
52	2	4	54	2	4
53	3	9	53	1	1
Total	7	37		7	23

$$\therefore \quad \bar{x} = 50 + \frac{7}{8} = \frac{407}{8}, \; \bar{y} = 52 + \frac{7}{8} = \frac{423}{8}$$

$$\Sigma(x-\bar{x})^2 = \Sigma(X-\bar{X})^2 = \Sigma X^2 - \frac{1}{n_1}(\Sigma X)^2 = 37 - \frac{49}{8} = \frac{247}{8}$$

$$\Sigma(y-\bar{y})^2 = \Sigma(Y-\bar{Y})^2 = 23 - \frac{49}{8} = \frac{135}{8}$$

$$\therefore \quad S^2 = \frac{\frac{247}{8} + \frac{135}{8}}{8+8-2} = \frac{382}{112}$$

$$\therefore \quad t = \frac{\frac{407}{8} \sim \frac{423}{8}}{\sqrt{\frac{382}{112}}\sqrt{\frac{1}{8}+\frac{1}{8}}} = 2{\cdot}17$$

$$\text{No. of } d.f. = 8 + 8 - 2 = 14$$

$$\therefore \quad t_{0{\cdot}05} = 2{\cdot}14 < t_{\text{cal}}$$

$\therefore$ At 5% level, difference between sample means is significant. Since $\bar{y} > \bar{x}$, food B is better.

(*b*) In this case the two samples cannot be regarded as independent. So the difference of means test for paired data will be applied.

									Total
x :	49	53	51	52	47	50	52	53	
y :	52	55	52	53	50	54	54	53	
$d=x-y$:	-3	-2	-1	-1	-3	-4	-2	0	$= -16$
d^2 :	9	4	1	1	9	16	4	0	$=$ 44

$$\bar{d} = -\frac{16}{8} = -2$$

$$\Sigma(d-\bar{d})^2 = 44 - \frac{1}{8}(-16)^2 = 12$$

$$\therefore \quad S^2 = \frac{12}{7}$$

$$\therefore \quad |t| = \frac{2\sqrt{8}}{\sqrt{\frac{12}{7}}} \simeq 4{\cdot}32$$

$$\text{No. of } d.f. = 8-1=7$$

$$\therefore \quad t_{0\cdot05} = 2{\cdot}36 \text{ and } t_{0\cdot01} = 3{\cdot}50$$

$$\therefore \quad |t_{cal}| > t_{0\cdot05} \text{ and } t_{0\cdot01}$$

$\therefore$ Difference between means is significant and hence food B is better.

EXERCISES

1. To compare the price of a certain commodity in two towns, ten shops were selected at random in each town. The following figures give the prices found :

Town A :	61	60	56	63	56	63	59	56	44	61
Town B :	55	54	47	59	51	61	57	54	62	58

Test whether the average price can be said to be the same in the two towns. **[Ans.** Yes]

2. Eight pots growing three wheat plants each were exposed to a high-tension discharge while nine similar pots were enclosed in an earthenware cage. The number of tillers in each pot were as follows :

Caged :	17	26	18	25	27	28	26	23	17
Electrified :	16	16	22	16	21	18	15	20	

Discuss whether electrification exercises any real effect on tillering. Given that $t_{0\cdot05}(15) = 2{\cdot}131$. **[Ans.** $|t| = 2{\cdot}75$]

3. In a test given to two groups of students the marks obtained were as follows :

First group :	18	20	36	50	49	36	34	41
Second group :	29	28	26	35	30	44	46	

Calculate student's t and state the relevant null hypothesis. **[Ans.** $|t| = 0{\cdot}28$]

4. The heights of six randomly chosen sailors are, in inches : 63, 65, 68, 69, 71 and 72. Those of ten randomly chosen soldiers are 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Discuss the light that these data throw on the suggestion that soldiers are on the average taller than sailors; given that

$$v = 14 \begin{cases} P = 0\cdot539 \text{ for } t = 0\cdot1 \\ P = 0\cdot527 \text{ for } t = 0\cdot08 \end{cases}$$

[Ans. $|t| = 0{\cdot}099$]

5. The means of two random samples of sizes 9 and 7 respectively are 196·42 and 198·82 respectively. The sum of the squares of the deviations from the means are 26·94 and 18·73 respectively. Can the samples be considered to have been drawn from the same

normal population, it bein

6. Two independent samples
Sample 1 : 9 1
Sample 2 : 10 1
Is the difference between t

v =

7. Two types of batteries A a
results are obtained :
No. in sample
A 10
B 10
Is there a significant differ

8. Intelligence test of two gro
if the difference of the mea
Group of 12 girls : n
Group of 8 boys : n

9. A farmer grows crops on tw
acre and on B Rs. 20 worth
on the two fields are :
Year
Yield A Rs. per acre
Yield B Rs. per acre
Other things beings equal,
to continue the more expen

10. Calculate the value of 't' in
frequencies are :
A : 16 10
B : 8 4

11. The yields of two types 'T
replications are given belo
the mean yields ? You may a
1·476.
Replication :
Yields in lbs : 2
(Type 17)
Yield in lbs : 2
(Type 51)

12. The following figures show
each was administered two
two)
Patient : 1 2
Soporific A : 1·2 1·8
Soporific B : 1·5 1·6
Apply an appropriate test to

normal population, it being given that $t_{0.05}(14) = 2.145$? **[Ans.** 2·6, No]

6. Two independent samples of 8 and 7 items respectively had the following values :

Sample 1 :	9	11	13	11	15	9	12	14
Sample 2 :	10	12	10	14	9	8	10	

Is the difference between the means of the samples significant ? Given that

$$v = 13 \begin{cases} P = 0.874 \text{ for } t = 1.2 \\ P = 0.892 \text{ for } t = 1.3 \end{cases}$$ **[Ans.** $|t| = 1.22$]

7. Two types of batteries A and B are tested for their length of life and the following results are obtained :

	No. in sample	*Mean*	*Variance*
A	10	500 hours	100
B	10	560 hours	121

Is there a significant difference in two means ? **[Ans.** Yes]

8. Intelligence test of two groups of boys and girls gave the following results. Examine if the difference of the means is significant.

Group of 12 girls : mean = 84, $s.d.$ = 10

Group of 8 boys : mean = 81, $s.d.$ = 12 **[Ans.** No]

9. A farmer grows crops on two fields A and B. On A he puts Rs. 10 worth of manure per acre and on B Rs. 20 worth. The net returns per acre, exclusive of the cost of manure, on the two fields are :

Year	1	2	3	4	5
Yield A Rs. per acre	34	28	42	37	44
Yield B Rs. per acre	36	33	48	38	50

Other things beings equal, discuss the question whether it is likely to pay the farmer to continue the more expensive dressing ? Given that $t_{0.05}(4) = 2.78$ **[Ans.** $t = 3.8$]

10. Calculate the value of 't' in case of two characteristics A and B whose corresponding frequencies are :

A :	16	10	8	9	9	8
B :	8	4	5	9	12	4

[Ans. $|t| = 1.66$]

11. The yields of two types 'Type 17' and 'Type 51' of grains in pounds per acre in a replications are given below. What comment would you make on the differences in the mean yields ? You may assume that if there be 5 d.f. and $P = 0.2$, t is approximately 1·476.

Replication :	1	2	3	4	5	6
Yields in lbs (Type 17) :	20·50	24·60	23·06	29·98	30·37	23·83
Yield in lbs (Type 51) :	24·86	26·39	28·19	30·75	29·97	22·04

[Ans. $t = 1.49$]

12. The following figures show the additional hours of sleep gained by 10 patients when each was administered two soporifics A and B (with an adequate period between the two)

Patient :	1	2	3	4	5	6	7	8	9	10
Soporific A :	1·2	1·8	– 0·3	– 0·7	0·1	3·1	2·2	– 1·5	0·0	2·1
Soporific B :	1·5	1·6	0·4	0·0	– 0·6	2·5	4·5	1·9	2·2	3·0

Apply an appropriate test to see whether the soporifics really differ in average effect.

[Ans. $t = 2.1$]

13. In each of 10 pairs of rats, one receives protein from raw peanut while the other receives it from roasted peanuts.

Pair	:	1	2	3	4	5	6	7	8	9	10
Raw	:	61	60	56	63	56	63	59	56	44	61
Roasted	:	55	54	47	69	51	61	57	54	62	58

Test whether or not roasting the peanuts had any effect on their protein value.

[**Ans.** Roasting had no effect on protein value]

14. In a rat-feeding experiment, the following results were obtained :

Diets	Gain in wts in gms											
High Protein:	13	14	10	11	12	16	10	9	11	12	9	12
Low Protein :	7	11	10	8	10	13	9					

Investigate if there is any evidence of superiority of one diet over the other.

[**Ans.** $|t| = 1{\cdot}93$]

15. Mitchel concluded a paired feeding experiment with pigs on the relative value of limestone and bonemeal for bone development. The results are given below :

Pair	:	1	2	3	4	5	6	7	8
Lime stone	:	49·2	53·3	50·6	52·0	46·8	50·5	52·1	53·0
Bonemeal	:	51·5	54·9	52·2	53·3	51·6	54·1	54·2	54·3

Determine the significance of the difference between the means in two ways (1) by assuming that the values are paired (2) by assuming that the values are not paired.

[**Ans.** Difference is significant]

16. Ten soldiers visit a rifle range for two consecutive weeks. For the first week their scores are :

67, 24, 57, 55, 63, 54, 56, 68, 33, 43

and during the second week they score, in the same order :

70, 38, 58, 58, 56, 67, 68, 72, 42, 38

Examine if there is any significant difference in their performance. [**Ans.** $|t| = 2{\cdot}04$]

17. Eleven school boys were given a test in mathematics. They were given a month's further tuition and a second test was held at the end of it. Do the marks give evidence that the students benefited by extra coaching ?

Roll No.		:	1	2	3	4	5	6	7	8	9	10	11
Marks	First test	:	23	20	19	21	18	20	18	17	23	16	19
	Second test	:	24	19	22	18	20	22	20	20	23	20	17

[**Ans.** $|t| = 1{\cdot}48$]

18. To test the desirability of a certain modification in typists desks, 9 typists were given two tests of as nearly as possible the same nature, one on the desk in use and the other on the new type. The following differences in the number of words typed per minute were recorded :

Typist	:	A	B	C	D	E	F	G	H	I
Increased no. of words per min	:	2	3	0	3	−1	4	−3	2	5

Do the data indicate that the modification in desk promoted speed in typing ?

[**Ans.** $t = 1{\cdot}96$]

19. Drugs A and B administered to 10 different persons each indicated the following rises in B.P.

A	:	3	0	–1	2	1	2	–1	0	1	3
B	:	– 1	– 3	0	1	1	0	–2	0	3	1

Stating the assumptions necessary, test the following null hypothesis :

(*i*) Drug A has no effect on B.P.

(*ii*) Drugs A and B are equally efficacious in respect of B.P.

20. Use t-test to examine the superiority of type I electric bulbs in the given data :

	Sample size	*Sample mean*	*Sample variance*
Type I	8	1234	1296
Type II	7	1036	1600

[**Ans.** $t = 9{\cdot}3$]

17.3-3. Test for the significance of an Observed Correlation Co-efficient

Consider random sample $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ from a bivariate normal population. Let the correlation co-efficient calculated from the sample be r and ρ the population correlation co-efficient. The hypothesis to be tested is :

'Whether population correlation co-efficient is zero *i.e.*, $\rho = 0$'.

Assuming this hypothesis, the statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

follows t-distribution with $(n-2)$ *d.f.**

To test the significance, calculated value of 't' is compared with the tabulated value and the significance is tested as usual.

Ex. 17-14. *Test whether the correlation is significant if* $r = 0{\cdot}6$, $n = 18$.

Sol.
$$t = \frac{(0\cdot 6)\sqrt{16}}{\sqrt{0\cdot 64}} = 3$$

$$\text{No. of } d.f. = 18 - 2 = 16$$

$$\therefore \quad t_{0\cdot 05} = 2\cdot 12 < t_{\text{cal}}$$

$\therefore$ Correlation is significant.

Ex. 17-15. *A random sample of* 18 *pairs from a bivariate normal population showed a correlation co-efficient of* 0·3. *Is this value significant of correlation in the population ?*

Sol.
$$t = \frac{(0\cdot 3)\sqrt{18-2}}{\sqrt{1-0\cdot 09}} = 1\cdot 26$$

$$\text{No. of } d.f. = 18 - 2 = 16$$

$$\therefore \quad t_{0\cdot 05} = 2\cdot 12 > t_{\text{cal}}$$

$\therefore$ Correlation is not significant.

Ex. 17-16. *Find the least value of 'r', in a sample of* 18 *pairs of observations from a bivariate normal population significant at 5% level.*

Sol.
$$\text{No. of } d.f., = 18 - 2 = 16$$

$$\therefore \quad t_{0\cdot 05} = 2\cdot 12$$

The least value of 'r' significant at 5% level is given by

$$\left|\frac{r\sqrt{18-2}}{\sqrt{1-r^2}}\right| > 2\cdot 12$$

*The proof of this is beyond the scope of this book.

:om raw peanut while the other

5 7 8 9 10
3 59 56 44 61
1 57 54 62 58

ect on their protein value.
g had no effect on protein value]
ere obtained :

10 9 11 12 9 12
9

'one diet over the other.
[**Ans.** $|t| = 1{\cdot}93$]

ith pigs on the relative value of
results are given below :

5 6 7 8
46·8 50·5 52·1 53·0
51·6 54·1 54·2 54·3

en the means in two ways (1) by
that the values are not paired.
[**Ans.** Difference is significant]

weeks. For the first week their

33, 43
order :
42, 38
ir performance. [**Ans.** $|t| = 2{\cdot}04$]

tics. They were given a month's
of it. Do the marks give evidence

6 7 8 9 10 11
20 18 17 23 16 19
22 20 20 23 20 17
[**Ans.** $|t| = 1{\cdot}48$]

ypists desks, 9 typists were given
ie on the desk in use and the other
umber of words typed per minute

E F G H I

–1 4 – 3 2 5
promoted speed in typing ?
[**Ans.** $t = 1{\cdot}96$]
each indicated the following rises

i.e., $16r^2 > (2{\cdot}12)^2(1-r^2) \;=\; 4{\cdot}4944(1-r^2)$

i.e., $20{\cdot}4944r^2 > 4{\cdot}4944$

i.e., $r^2 > 0{\cdot}2193$

i.e., $|r| > 0{\cdot}4683$

$\therefore$ Regd. value of $|r| = 0{\cdot}4683$.

Ex. 17-17. *A random sample of* 15 *from a normal population gives a correlation co-efficient of* $-0{\cdot}5$. *Is this significant of the existence of correlation in the population ?*

Sol.
$$t = \frac{(-0{\cdot}5)\sqrt{13}}{\sqrt{0{\cdot}75}} = -2{\cdot}08$$

$t_{0{\cdot}05}$ for 13 *d.f.* $= 2{\cdot}16 > |t_{\text{cal}}|$

$\therefore$ Sample correlation co-efficient is not significant.

17.3-4. Test for the significance of an observed Regression Co-efficient

Consider a random sample (x_1, y_1) ..... (x_n, y_n) from a bivariate normal population. Let the equation of line of regression of y on x (obtained from the sample) be

$$y-\bar{y} = b(x-\bar{x})$$

where b = regression co-eff. of y on x.

Let
$$Y_i = \bar{y}+b(x_i-\bar{x}).$$

The hypothesis to be tested is : "The regression co-efficient of y on x in the population is β". Assuming this hypothesis, the statistic

$$t = (b-\beta)\left\{\frac{(n-2)\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(y_i-Y_i)^2}\right\}^{1/2}$$

follows t-distribution with $(n-2)$ d.f.

17.3-5. Test for the Significance of a Rank Correlation Co-efficient

Let ρ be the rank correlation co-efficient obtained from a sample of size n. The hypothesis to be tested is :

"population rank correlation coefficient is zero."

Assuming this hypothesis, the statistic

$$t = \rho\left\{\frac{n-2}{1-\rho^2}\right\}^{1/2}$$

follows t-distribution with $(n-2)$ d.f.

Ex. 17-18. *12 pictures submitted in a competition were ranked by two judges with results as shown in the table below :*

Pictures	:	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Rank assigned by 1st judge	:	5	9	6	7	1	3	4	12	2	11	10	8
Rank assigned by 2nd judge	:	5	8	9	11	3	1	2	10	4	12	7	6

Calculate ρ *the rank correlation co-efficient. Is there a lack of independence in these ranking ? (Assume that on the hypothesis of independence of two sets of n readings*

$t = \rho\left(\frac{n-2}{1-\rho^2}\right)^{1/2}$ *follows t-dis*

Sol. Let d be the differenc

Now $\Sigma d^2 = 0+1+9+16$

$\therefore$ ρ

$\therefore$ t

No. of *d.f.*

$\therefore$ $t_{0{\cdot}05}$

$\therefore$ Ranking is significan

1. Is a correlation co-efficien pairs of values from a nor
2. Fine the least value of r, significant at 5% level. (G
3. Determine the range withi in random sampling from (Given that $P\{|t(8)| > 2{\cdot}3$
4. Determine the least value random sampling from a b
   (*i*) the sample size is 5,
   (*ii*) the sample size is mo

17.4. F-distribution

Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2,$ be two independent random sam σ^2.

Let $\bar{x}$ and $\bar{y}$ be the sample

$S_1^2 =$

The statistic F is defined by

$F =$

Let $\nu_1 =$

$\therefore$ $\frac{\nu_1 F}{\nu_2} =$

$=$

$t = \rho\left(\dfrac{n-2}{1-\rho^2}\right)^{1/2}$ *follows t-distribution with* $(n-2)$ *d.f. Given that for* 10 *d.f.*, $t_{0\cdot05} = 2\cdot23$.

Sol. Let d be the difference in ranks assigned to the same individuals.

Now $\Sigma d^2 = 0 + 1 + 9 + 16 + 4 + 4 + 4 + 4 + 4 + 1 + 9 + 4 = 60$

$$\therefore \quad \rho = 1 - \frac{(6)(60)}{(12)(143)} \simeq 0\cdot79$$

$$\therefore \quad t = (0\cdot79)\frac{\sqrt{10}}{\sqrt{1-(0\cdot79)^2}} = \frac{(0\cdot79)\sqrt{10}}{\sqrt{0\cdot3759}}$$

$$\simeq 4\cdot075$$

$$\text{No. of } d.f. = 12 - 2 = 10$$

$$\therefore \quad t_{0\cdot05} = 2\cdot23 < t_{cal}$$

$\therefore$ Ranking is significant and hence there is lack of independence.

EXERCISES

1. Is a correlation co-efficient of 0·5 significant, if obtained from a random sample of 12 pairs of values from a normal population ? [**Ans.** No.]
2. Fine the least value of r, in a sample of 25 pairs from normal population, which is significant at 5% level. (Given that for 23 $d.f.$ $t_{0\cdot05} = 2\cdot07$). [**Ans.** 0·4]
3. Determine the range within which r will not be significant at 5% level of significance in random sampling from a bivariate normal population when the sample size is 10. (Given that $P\{|t(8)| > 2\cdot306\} = 0\cdot05$)
4. Determine the least value of $|r|$ that will be significant at 5% level of significance in random sampling from a bivariate normal population when
   (*i*) the sample size is 5,
   (*ii*) the sample size is more than 30.
   (Given that $P\{|t(3)| > 3\cdot182\} = 0\cdot05$)

17.4. F-distribution

Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2, \ldots y_{n_2}$ be two independent random samples drawn from the same normal population with variance σ^2.

Let $\bar{x}$ and $\bar{y}$ be the sample means and

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2,\; S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_1}(y_i-\bar{y})^2$$

The statistic F is defined by

$$F = \frac{S_1^2}{S_2^2} \quad \left(S_1^2 > S_2^2\right)$$

Let $\quad \nu_1 = n_1 - 1, \quad \nu_2 = n_2 - 1$

$$\therefore \quad \frac{\nu_1 F}{\nu_2} = \left(\frac{(n_1-1)S_1^2}{(n_2-1)S_2^2}\right)$$

$$= (n_1 s_1^2/\sigma^2)\,/\,(n_2 s_2^2/\sigma^2)$$

Now $n_1 s_1^2/\sigma^2$ is a ψ^2 variate with ν_1 *d.f.* and $n_2 s_2^2/\sigma^2$ is a ψ^2 variate with ν_2 *d.f.*

$\therefore \quad \dfrac{\nu_1 F}{\nu_2}$ is a $\beta_2\left(\dfrac{\nu_1}{2}, \dfrac{\nu_2}{2}\right)$ variate.

$\therefore$ Distribution of F is

$$dP = \frac{1}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \frac{\left(\frac{\nu_1 F}{\nu_2}\right)^{\frac{\nu_1}{2}-1}}{\left(1+\frac{\nu_1 F}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}} d\left(\frac{\nu_1 F}{\nu_2}\right)$$

$$= \frac{1}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}} \cdot F^{\frac{\nu_1}{2}-1}}{(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}}} dF \quad 0 < F < \infty$$

This distribution is called the distribution of the variance ratio F with ν_1 and ν_2 *d.f.*

Ex. *Let ψ_1^2 and ψ_2^2 be two independent chi-square variates with n_1 and n_2 d.f. Find the distribution of* $F = \dfrac{\psi_1^2/n_1}{\psi_2^2/n_2}$.

17.4-1. Constants of F-Distribution

$$\mu'_r(0) = E(F^r)$$

$$= \frac{\nu_1^{\frac{\nu_1}{2}} \cdot \nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \int_0^\infty F^r \frac{F^{\frac{\nu_1}{2}-1}}{(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}}} dF$$

$$= \frac{\nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \int_0^\infty \frac{F^{\frac{\nu_1}{2}+r-1}}{(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}}} dF$$

Put $\quad \nu_1 F = \nu_2 x \Rightarrow dF = \dfrac{\nu_2\, dx}{\nu_1}$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \frac{1}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \int_0^\infty \frac{x^{\frac{\nu_1}{2}+r-1}}{(1+x)^{\left(\frac{\nu_1}{2}+r\right)+\left(\frac{\nu_2}{2}-r\right)}} dx$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \frac{1}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \beta\left(\frac{\nu_1}{2}+r, \frac{\nu_2}{2}-r\right)$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \frac{\Gamma\left(\frac{\nu_1}{2}+r\right)\Gamma\left(\frac{\nu_2}{2}-r\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}$$

$\therefore$ Mean =

Put r =

$\mu'_2(0)$ =

$\therefore$ $\mu_2 = \mu$

Mode. The density function of

$f(F) =$

$\dfrac{df}{dF} =$

$\dfrac{df}{dF} =$

is a ψ^2 variate with ν_2 *d.f.*

nce ratio F with ν_1 and ν_2 *d.f.*
variates with n_1 and n_2 d.f. Find

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \frac{\left(\frac{\nu_1}{2}+r-1\right).....\left(\frac{\nu_1}{2}\right)}{\left(\frac{\nu_2}{2}-1\right).....\left(\frac{\nu_2}{2}-r\right)}$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \frac{\{\nu_1+2(r-1)\}....(\nu_1)}{(\nu_2-2)....(\nu_2-2r)}$$

$$\therefore \quad \text{Mean} = \mu_1'(0) = \frac{\nu_2}{\nu_2-2} > 1.$$

Put $r = 2$

$$\mu'_2(0) = \left(\frac{\nu_2}{\nu_1}\right)^2 \frac{\nu_1(\nu_1+2)}{(\nu_2-2)(\nu_2-4)}$$

$$= \frac{\nu_2^2(\nu_1+2)}{\nu_1(\nu_2-2)(\nu_2-4)}$$

$$\therefore \quad \mu_2 = \mu'_2(0) - \{\mu'_1(0)\}^2$$

$$= \frac{\nu_2^2(\nu_1+2)}{\nu_1(\nu_2-2)(\nu_2-4)} - \left(\frac{\nu_2}{\nu_2-2}\right)^2$$

$$= \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$$

Mode. The density function of F variate is

$$f(F) = \frac{1}{\beta\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \frac{\nu_1^{\frac{\nu_1}{2}}\,\nu_2^{\frac{\nu_2}{2}}\,F^{\frac{\nu_1}{2}-1}}{(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}}}$$

$$\frac{df}{dF} = \frac{\nu_1^{\frac{\nu_1}{2}}\,\nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \left\{\frac{\left(\frac{\nu_1}{2}-1\right)F^{\frac{\nu_1}{2}-2}(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}} - F^{\frac{\nu_1}{2}-1}\left(\frac{\nu_1+\nu_2}{2}\right)(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}-1}_{\nu_1}}{(\nu_2+\nu_1 F)^{\nu_1+\nu_2}}\right\}$$

$$\frac{df}{dF} = \frac{\nu_1^{\frac{\nu_1}{2}}\,\nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} F^{\frac{\nu_1}{2}-2}(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}-1}.$$

$$\left\{\frac{\left(\frac{\nu_1}{2}-1\right)(\nu_2+\nu_1 F) - F.\left(\frac{\nu_1+\nu_2}{2}\right)\nu_1}{(\nu_2+\nu_1 F)^{\nu_1+\nu_2}}\right\}$$

$$\frac{df}{dF} = 0 \Rightarrow$$

$$F = 0, \text{ and}$$

$$\left(\frac{\nu_1}{2} - 1\right)(\nu_2 + \nu_1 F) - F.\left(\frac{\nu_1 + \nu_2}{2}\right)\nu_1 = 0$$

$$\Rightarrow \qquad F = \frac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)}$$

for $F = 0$, $f(F) = 0$ which is minimum value of $f(F)$.

$\therefore$ for $F = \dfrac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)}$, $f(F)$ is maximum

$$\therefore \qquad \text{Mode} = \frac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)}.$$

17.4.2. Chief features of F-Probability Cure

The equation of the F-probability curve is

$$y = \frac{\nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2}; \frac{\nu_2}{2}\right)} \frac{F^{\frac{\nu_1}{2}-1}}{(\nu_2 + \nu_1 F)^{\frac{\nu_1+\nu_2}{2}}}. \quad 0 < F < \infty.$$

(1) At $F = 0$, $y = 0$

and as $F \to \infty$, $y \to 0$

$\therefore$ F-axis is asymptote to the curve at positive extremity.

(2) Mode is at the point

$$F = \frac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)}$$

which exists only when $\nu_1 > 2$ $(\because F \geq 0)$

Now

$$\text{Mode} - 1 = \frac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)} - 1$$

$$= \frac{-2(\nu_1 + \nu_2)}{\nu_1(\nu_2 + 2)} < 0$$

$$\Rightarrow \qquad \text{Mode} < 1.$$

(3) Karl Pearson's co-efficient of skewness is

$$\frac{\text{mean} - \text{mode}}{s.d.}$$

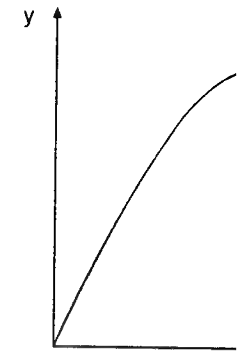
Since mean > 1 and mode < 1.

F probability curve is highly positively skewed.

(4) The pts of inflextion of F curve exist when $\nu_1 > 4$ and are equidistant from mode (See 17.4.3).

(5) y increases steadily at first until it reaches is maximum value and then decreases slowly.

$\therefore$ The shape of the probab



17.4.3. Point of Inflextion

The equation of the probabilit

$$y = \frac{\nu_1}{\beta(}$$

Put $\nu_1 F = \nu_2 x$

Then $y = c\,.$

where $c = \dfrac{\nu_1}{\nu_2}$

$\therefore$ $\log y = \log$

Differentiating w.r.t. x

$$\frac{1}{y}\frac{dy}{dx} = (l -$$

Differentiating again

$$-\frac{1}{y^2}\left(\frac{dy}{dx}\right)^2 + \frac{1}{y}\frac{d^2y}{dx^2} = \frac{-(l}{x}$$

At points of inflextion

$$\frac{d^2y}{dF^2} = \left(\frac{\nu_1}{\nu_2}\right.$$

and $\dfrac{d^3y}{dF^3} = \left(\dfrac{\nu_1}{\nu_2}\right.$

$$\Rightarrow \qquad \frac{d^2y}{dx^2} = 0,$$

(F).

imum

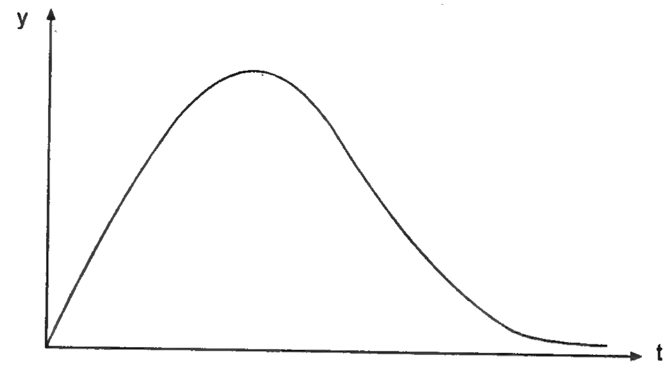
$\frac{\nu_1+\nu_2}{2}$. $0 < F < \infty$.

tremity.

> 4 and are equidistant from mode

naximum value and then decreases

$\therefore$ The shape of the probability curve is approximately as shown.



17.4.3. Point of Inflextion

The equation of the probability curve is

$$y = \frac{\nu_1^{\frac{\nu_1}{2}}\,\nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}\;\frac{F^{\frac{\nu_1}{2}-1}}{(\nu_2+\nu_1 F)^{\frac{\nu_1+\nu_2}{2}}}$$

Put $\qquad \nu_1 F = \nu_2 x.$

Then
$$y = c\,.\,\frac{x^{l-1}}{(1+x)^{l+m}}$$

where
$$c = \frac{\nu_1}{\nu_2}\,\frac{1}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)},\; l = \frac{\nu_1}{2},\; m = \frac{\nu_2}{2}$$

$$\therefore \qquad \log y = \log c + (l-1)\log x - (l+m)\log(1+x)$$

Differentiating w.r.t. x

$$\frac{1}{y}\frac{dy}{dx} = (l-1)\frac{1}{x} - (l+m)\,\frac{1}{1+x}$$

Differentiating again

$$-\frac{1}{y^2}\left(\frac{dy}{dx}\right)^2 + \frac{1}{y}\frac{d^2y}{dx^2} = \frac{-(l-1)}{x^2} + \frac{l+m}{(1+x)^2}$$

At points of inflextion

$$\frac{d^2y}{dF^2} = \left(\frac{\nu_1}{\nu_2}\right)^2 \frac{d^2y}{dx^2} = 0$$

and
$$\frac{d^3y}{dF^3} = \left(\frac{\nu_1}{\nu_2}\right)^3 \frac{d^3y}{dx^3} \neq 0$$

$$\Rightarrow \qquad \frac{d^2y}{dx^2} = 0,\; \frac{d^3y}{dx^3} \neq 0$$

$\therefore$ Points of inflextion are given by

$$\therefore \quad \frac{-1}{y^2}\left(\frac{dy}{dx}\right)^2 = -\frac{(l-1)}{x^2}+\frac{l+m}{(1+x)^2}$$

i.e.,
$$\left\{(l-1)\frac{1}{x}-\frac{l+m}{1+x}\right\}^2-\frac{l-1}{x^2}+\frac{l+m}{(1+x)^2}=0$$

i.e.,
$$\{(l-1)(1+x)-(l+m)x\}^2-(l-1)(1+x)^2+(l+m)x^2=0$$

i.e.,
$$\{(l-1)^2-(l-1)\}(1+x)^2+\{(l+m)^2+(l+m)\}x^2-2(l-1)(l+m)x(1+x)=0$$

i.e.,
$$x^2\{(l-1)^2-(l-1)+(l+m)^2+(l+m)-2(l-1)(l+m)\}$$
$$+\ 2x\{(l-1)^2-(l-1)-(l-1)(l+m)\}+\{(l-1)^2-(l-1)\}=0$$

The roots of this equation give two points of inflextion. Let these be x_1 and x_2

$$\therefore \quad x_1+x_2 = \frac{-2\{(l-1)^2-(l-1)-(l-1)(l+m)\}}{(l-1)^2-(l-1)+(l+m)^2+(l+m)-2(l-1)(l+m)}$$

$$= \frac{2(l-1)(m+2)}{-(l-1)(m+2)+(l+m)(m+2)}$$

$$= \frac{2(l-1)}{m+1}$$

$$\therefore \quad x_1+x_2 = \frac{2(\nu_1-2)}{\nu_2+2}$$

Let F_1 and F_2 be the corresponding values of F. Then F_1 and F_2 are pts. of inflextion.

We have
$$\nu_1F_1 = \nu_2x_1 \Rightarrow F_1 = \frac{\nu_2}{\nu_1}x_1$$

and
$$F_2 = \frac{\nu_2}{\nu_1}x_2$$

$$\therefore \quad F_1+F_2 = \frac{\nu_2}{\nu_1}(x_1+x_2)$$

$$= \frac{\nu_2}{\nu_1}\cdot\frac{2(\nu_1-2)}{\nu_2+2}$$

$$= 2\ (\text{mode}).$$

$\therefore$ F_1 — mode = mode — F_2.

$\therefore$ F_1 and F_2 are equidistant from mode

The condition $\dfrac{d^3y}{dx^3} \neq 0 \Rightarrow \nu_1 > 4.$*

$\therefore$ Points of infflextion exist if $\nu_1 > 4$.

Ex. 17-19. (*a*) *If x has a F-distribution with (m, n) d.f. show that $\dfrac{1}{x}$ has a F-distribution with (n, m) d.f.*

(*b*) *Deduce that, for any $k > 0$*

$$P\{x \le k\} + P\left\{y \le \frac{1}{k}\right\} = 1$$

*The proof of this is left as an exercise for the reader.

where x and y are F-distributed

Sol. (*a*) Dist. of x is

$dP = -$

β

Put $x = \dfrac{1}{y} \Rightarrow dx = -\dfrac{1}{y^2}$

$\therefore$ Dist. of y is

$dP = -$

β

$= -$

β

$\Rightarrow \quad y = \dfrac{1}{x}$ is F distribute

(*b*) Since total prob. is unity,

$P\{x$

Now $\quad P\{x \ge k\} = P$

$= P$

where $\quad y = \dfrac{1}{x}$ is F-distribute

$\therefore \quad P\{x \le k\} + P\left\{y \le \dfrac{1}{k}\right\}$

Ex. 17-20. *If $\nu_1 = \nu_2$, the me*

$Q_1Q_3 = 1$

where Q_1, Q_3 are quartiles.

Sol. Let $\nu_1 = \nu_2 = \nu$.

Let x be F distributed with (ν

Then $y = \dfrac{1}{x}$ also is F-distr

$\therefore \quad P\{x \le a\} = P$

(*i*) Let k be the median.

Then $\quad P\{x \le k\} = P$

$= P$

$+ (l+m)\, x^2 = 0$

$)\}\, x^2 - 2(l-1)(l+m)\, x\, (1+x) = 0$

$-1)(l+m)\}$

$l-1)^2 - (l-1)\} = 0$

tion. Let these be x_1 and x_2

$\dfrac{1)-(l-1)(l+m)\}}{+(l+m)-2(l-1)(l+m)}$

$(m+2)$

hen F_1 and F_2 are pts. of inflextion.

c_1

1) d.f. show that $\dfrac{1}{x}$ *has a F-distribution*

where x and y are F-distributed with (m, n) and (n, m) d.f.s. respectively.

Sol. (*a*) Dist. of x is

$$dP = \frac{m^{\frac{m}{2}}\, n^{\frac{n}{2}}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}}\, dx,\ 0 \le x < \infty$$

Put $x = \dfrac{1}{y} \Rightarrow dx = -\dfrac{1}{y^2}\, dy$

$\therefore$ Dist. of y is

$$dP = \frac{m^{\frac{m}{2}}\, n^{\frac{n}{2}}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{\left(\frac{1}{y}\right)^{\frac{m}{2}-1}}{\left(n+\frac{m}{y}\right)^{\frac{m+n}{2}}} \frac{1}{y^2}\, dy$$

$$= \frac{m^{\frac{m}{2}}\, n^{\frac{n}{2}}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{y^{\frac{n}{2}-1}}{(m+ny)^{\frac{m+n}{2}}}\, dy \qquad 0 \le y < \infty$$

$\Rightarrow$ $y = \dfrac{1}{x}$ is F distributed with (n, m) d.f.

(*b*) Since total prob. is unity,

$$P\{x \le k\} + P\{x \ge k\} = 1$$

Now $$P\{x \ge k\} = P\left\{\frac{1}{x} \le \frac{1}{k}\right\} = P\left\{y \le \frac{1}{k}\right\}$$

where $y = \dfrac{1}{x}$ is F-distributed with (n, m) d.f.

$$\therefore\quad P\{x \le k\} + P\left\{y \le \frac{1}{k}\right\} = 1.$$

Ex. 17-20. *If* $\nu_1 = \nu_2$, *the median of F distribution is at* $F = 1$. *Show also that*

$$Q_1 Q_3 = 1$$

where Q_1, Q_3 *are quartiles.*

Sol. Let $\nu_1 = \nu_2 = \nu$.

Let x be F distributed with (ν, ν) d.f.

Then $y = \dfrac{1}{x}$ also is F-distributed with (ν, ν) d.f.

$\therefore$ $P\{x \le a\} = P\{y \le a\}$, for any a.

(*i*) Let k be the median.

Then $$P\{x \le k\} = P\{x \ge k\} = P\left\{y \le \frac{1}{k}\right\}$$

$$= P\left\{x \le \frac{1}{k}\right\}$$

which is possible only when $k = 1$.

$\therefore$ Median $= 1$.

(*ii*) Now $P\{x \le Q_1\} = P\{x \ge Q_3\}$

$$= P\left(y \le \frac{1}{Q_3}\right)$$

$$= P\left(x \le \frac{1}{Q_3}\right)$$

which is possible only when

$$Q_1 = \frac{1}{Q_3}$$

i.e., $$Q_1 Q_3 = 1$$

EXERCISES

1. If $\nu_1 = 2$, show that

$$P(F \ge F_0) = \left(1 + \frac{2F_0}{\nu_2}\right)^{-\frac{\nu_2}{2}}$$

2. If x has a F distribution with (n_1, n_2) *d.f.*, show that

$$\left(1 + n_1 \frac{x}{n_2}\right)^{-1}$$

has a Beta distribution.

3. If $\nu_1 = \nu_2 = n - 1$, show that

$$\text{H.M.} = \frac{n-1}{n-3}.$$

4. Given that

$$P\{F(10, 12) > 2 \cdot 753\} = 0 \cdot 05$$

and $$P\{F(1, 12) > 4 \cdot 747\} = 0 \cdot 05$$

find

(*i*) $P\left\{F(12, 10) > \dfrac{1}{2 \cdot 753}\right\}$

(*ii*) $P\left\{-\sqrt{4 \cdot 747} < t(12) < \sqrt{4 \cdot 747}\right\}$.

17.4-4. Relation between t and F distributions

Student's t-distribution is

$$dP = \frac{1}{\sqrt{\nu}} \frac{1}{\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \frac{dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}, \quad -\infty < t < \infty$$

where ν = No. of *d.f.*

Put $\quad t^2 = x \Rightarrow t = \sqrt{x}$

$$\therefore \quad dt = \frac{1}{2\sqrt{x}}\, dx$$

$$\therefore \quad dP = 2 \cdot \frac{1}{2\sqrt{v}} \frac{1}{\beta\left(\frac{v}{2}, \frac{1}{2}\right)} \frac{x^{-\frac{1}{2}}\, dx}{\left(1 + \frac{x}{v}\right)^{\frac{v+1}{2}}}$$

$$= \frac{1}{\sqrt{v}} \cdot \frac{1}{\beta\left(\frac{v}{2}, \frac{1}{2}\right)} \cdot \frac{x^{\frac{1}{2}-1}}{\left(1 + \frac{x}{v}\right)^{\frac{v+1}{2}}}\, dx, \quad 0 < x < \infty$$

$$= \frac{v^{\frac{v}{2}}}{\beta\left(\frac{v}{2}, \frac{1}{2}\right)} \frac{x^{\frac{1}{2}-1}}{(v + x)^{\frac{v+1}{2}}}\, dx$$

which $\Rightarrow$ x is a F variate with d.f. 1 and v.

Remark. All tests of significance basesd on t-distribution can be done by using F-distribution.

17.4-5. Relation between F and ψ_4^2

F-distribution with v_1, v_2 d.f. is

$$dP = \frac{1}{\beta\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} v_1^{\frac{v_1}{2}} v_2^{\frac{v_2}{2}} \frac{F^{\frac{v_1}{2}-1}}{(v_2 + v_1 F)^{\frac{v_1+v_2}{2}}}\, dF$$

$$= \left\{\frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_2}{2}\right)} \frac{1}{v_2^{\frac{v_1}{2}}}\right\} \frac{1}{\Gamma\left(\frac{v_1}{2}\right)} \frac{v_1^{\frac{v_1}{2}} F^{\frac{v_1}{2}-1}}{\left(1+\frac{v_1 F}{v_2}\right)^{\frac{v_1+v_2}{2}}}\, dF$$

Let $$\Lambda = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_2}{2}\right)} \cdot \frac{1}{v_2^{\frac{v_1}{2}}}$$

$$= \frac{\sqrt{2\pi}\, e^{-\left\{\frac{v_1+v_2}{2}-1\right\}}}{\sqrt{2\pi}\, e^{-\left\{\frac{v_2}{2}-1\right\}}} \cdot \frac{\left\{\frac{v_1+v_2}{2}-1\right\}^{\frac{v_1+v_2}{2}-\frac{1}{2}}}{\left\{\frac{v_2}{2}-1\right\}^{\frac{v_2}{2}-\frac{1}{2}}} \cdot \frac{1}{v_2^{\frac{v_1}{2}}}$$

$$= \frac{e^{-\frac{\nu_1}{2}}}{2^{\frac{\nu_1}{2}}} \frac{(\nu_1+\nu_2-2)^{\frac{\nu_1+\nu_2-1}{2}}}{(\nu_2-2)^{\frac{\nu_2}{2}-\frac{1}{2}}} \cdot \frac{1}{\nu_2^{\frac{\nu_1}{2}}}$$

$$= \frac{e^{-\frac{\nu_1}{2}} \cdot 2^{-\frac{\nu_1}{2}} \left\{1+\frac{\nu_1-2}{\nu_2}\right\}^{\frac{\nu_1+\nu_2-1}{2}}}{\left(1-\frac{2}{\nu_2}\right)^{\frac{\nu_2}{2}-\frac{1}{2}}}$$

$$= \frac{e^{-\frac{\nu_1}{2}} \cdot 2^{-\frac{\nu_1}{2}} \left(1+\frac{\nu_1-2}{\nu_2}\right)^{\frac{\nu_1-1}{2}} \left[\left\{\left(1+\frac{\nu_1-2}{\nu_2}\right)\right\}^{\frac{\nu_2}{\nu_1-2}}\right]^{\frac{\nu_1-2}{2}}}{\left\{\left(1-\frac{2}{\nu_2}\right)^{-\frac{\nu_2}{2}}\right\}^{-1} \cdot \left\{1-\frac{2}{\nu_2}\right\}^{-\frac{1}{2}}}$$

$$\to e^{-\frac{\nu_1}{2}} \cdot 2^{-\frac{\nu_1}{2}} \cdot \frac{e^{\frac{\nu_1}{2}-1}}{e^{-1}} = 2^{-\frac{\nu_1}{2}}$$

as $\nu_2 \to \infty$

Also $$\left(1+\frac{\nu_1 F}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}} = \left[\left\{1+\frac{\nu_1 F}{\nu_2}\right\}^{\frac{\nu_2}{\nu_1 F}}\right]^{\frac{\nu_1 F}{2}} \left(1+\frac{\nu_1 F}{\nu_2}\right)^{\frac{\nu_1}{2}}$$

$$\to e^{\frac{\nu_1 F}{2}} \text{ as } \nu_2 \to \infty$$

$\therefore$ As $\nu_2 \to \infty$, probability differential is of the form

$$dP = \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)} 2^{-\frac{\nu_1}{2}} \nu_1^{\frac{\nu_1}{2}} F^{\frac{\nu_1}{2}-1} e^{-\frac{\nu_1 F}{2}} dF$$

$$= \frac{1}{2^{\frac{\nu_1}{2}} \Gamma\left(\frac{\nu_1}{2}\right)} e^{-\frac{\nu_1 F}{2}} (\nu_1 F)^{\frac{\nu_1}{2}-1} d(\nu_1 F)$$

$\Rightarrow$ $\nu_1 F$ is a ψ^2 – variate with ν_1 *d.f.*

17.5. F-tests

Tests of significance based of F-distribution are called F tests. Various F-tests are :

(*i*) For equality of population variance.

(*ii*) For the significance of an observed multiple correlation co-efficient.

(*iii*) For the significance of an observed sample correlation ratio.

(*iv*) For testing the linearity of regression.

All these tests are for small samples.

Rules of Decision

Let $P = P\{F > F_0(\nu_1, \nu_2)\}$

For a given value of P and f

of F-tables.

The value F_0 is called the cr

To test the significance the

certain specified level of signific

If $F_{cal} > F_{tab}$, the null hypot

and if $F_{cal} < F_{tab}$, the hypothesis

Ex. 17-21. *When* ν_1 = 2, *s*

significant probability p is

F =

where ν_1 *and* ν_2 *have their usual*

Sol. The dist. of F *is*

dP = – β

Let F_0 be the significance le

Then $p = P$

= – β

= – β

= 2 β

= 2

= 2

F tests. Various F-tests are :

elation co-efficient.
ation ratio.

Rules of Decision

Let $P = P\{F > F_0(\nu_1, \nu_2)\}$

For a given value of P and for ν_1, ν_2 *d.f.s*, values of F_0 have been tabulated in the form of F-tables.

The value F_0 is called the critical value of F for ν_1, ν_2 *d.f.s.* at level of significance P.

To test the significance the calculated value of F is compared with tabulated value at certain specified level of significance. Generally 5% or 1% levels are taken.

If $F_{cal} > F_{tab}$, the null hypothesis is rejected and the difference is said to be significant and if $F_{cal} < F_{tab}$, the hypothesis is accepted at the level of significance adopted.

Ex. 17-21. *When $\nu_1 = 2$, show that the significance level of F corresponding to a significant probability p is*

$$F = \frac{\nu_2}{2}\left(p^{\frac{-2}{\nu_2}} - 1\right)$$

where ν_1 and ν_2 have their usual meanings.

Sol. The dist. of F is

$$dP = \frac{2 \cdot \nu_2^{\frac{\nu_2}{2}}}{\beta\left(1, \frac{\nu_2}{2}\right)} \cdot \frac{dF}{(\nu_2 + 2F)^{\frac{\nu_2}{2}+1}}$$

Let F_0 be the significance level of F corresponding to the probability p.

Then
$$p = P\{F \geq F_0\} = \int_{F_0}^{\infty} dP$$

$$= \frac{2\nu_2^{\frac{\nu_2}{2}}}{\beta\left(1, \frac{\nu_2}{2}\right)} \int_{F_0}^{\infty} \frac{dF}{(\nu_2 + 2F)^{\frac{\nu_2}{2}+1}}$$

$$= \frac{2\nu_2^{\frac{\nu_2}{2}}}{\beta\left(1, \frac{\nu_2}{2}\right)} \left\{-\frac{1}{\nu_2} \cdot \frac{1}{(\nu_2 + 2F)^{\frac{\nu_2}{2}}}\right\}_{F_0}^{\infty}$$

$$= \frac{2\nu_2^{\frac{\nu_2}{2}-1}}{\beta\left(1, \frac{\nu_2}{2}\right)} \frac{1}{(\nu_2 + 2F_0)^{\frac{\nu_2}{2}}}$$

$$= 2\nu_2^{\frac{\nu_2}{2}-1} \frac{\Gamma\left(\frac{\nu_2}{2}+1\right)}{\Gamma(1)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \frac{1}{(\nu_2 + 2F_0)^{\frac{\nu_2}{2}}}$$

$$= 2\nu_2^{\frac{\nu_2}{2}-1} \frac{\frac{\nu_2}{2}\Gamma\left(\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \frac{1}{(\nu_2 + 2F_0)^{\frac{\nu_2}{2}}} \qquad (\because \Gamma(1) = 1)$$

$$= \frac{\nu_2^{\frac{\nu_2}{2}}}{(\nu_2 + 2F_0)^{\frac{\nu_2}{2}}} \Rightarrow F_0 = \frac{\nu_2}{2}\left(p^{-\frac{2}{\nu_2}} - 1\right).$$

17.5.1. Test of significance for equality of population variance

Consider two independent random samples $x_1, x_2, \ldots\ldots x_{n_1}$ and $y_1, y_2, \ldots\ldots y_{n_2}$ from normal populations. The hypothesis to be tested is: 'The population variances are same'.

Assuming this hypothesis, the statistic

$$F = \frac{S_1^2}{S_2^2} \quad (S_1^2 > S_2^2)$$

where $$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$$

and $$S_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(y_i - \bar{y})^2,$$

follows F-distribution with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom.

By comparing the calculated value of F with the tabulated value for ν_1 and ν_2 *d.f.* at certain level of significance (5% or 1%) the significance is tested.

Ex. 17-22. *If is known that the mean diameters of rivets produced by two firms A and B are practically the same but the standard deviations may differ. For 22 rivets produced by A the s.d. is 2·9 m, while for 16 rivets manufactured by B the s.d. is 3·8. Test whether the products of A have the same variability as those of B.*

Sol. $n_1 = 22,\ n_2 = 16,\ s_1 = 2{\cdot}9$ and $s_2 = 3{\cdot}8$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{(22)(2{\cdot}9)^2}{21} \simeq 8{\cdot}81$$

and $$S_2^2 = \frac{(16)(3{\cdot}8)^2}{15} \simeq 15{\cdot}40$$

$$\therefore \quad F = \frac{S_2^2}{S_1^2} = \frac{15{\cdot}4}{8{\cdot}81} \simeq 1{\cdot}748$$

Nos. of degrees of freedom are $16 - 1 = 15$ and $22 - 1 = 21$

$\therefore \quad F_{0\cdot05} = 2{\cdot}18 > F_{\text{cal}}$

$\therefore$ Variability for two types of products may be same.

Ex. 17-23. *Given below are the qualities of ten items (in proper units) produced by two processes A and B. Test whether the variability of quality may be taken to be the same for the two processes.*

Process A :	3	7	5	6	5	4	4	5	3	3
Process B :	8	5	7	8	3	2	7	6	5	7

(F-value for $n_1 = 9$ and $n_2 = 9$ *degrees of freedom is* 3·18 *at* 5% *level of significance and* 5·35 *at* 1% *level of significance*).

Sol.

Process $A : x \rightarrow$	3	7	5	6	5	4	Total
$X = (x - \bar{x}) \rightarrow$	$-1{\cdot}5$	2·5	0·5	1·5	0·5	$-0{\cdot}5$	
$X^2 \rightarrow$	2·25	6·25	0·25	2·25	0·25	0·25	

Process $B : y \rightarrow$ 8
$Y = (y - \bar{y}) \rightarrow$ 2·2
$Y^2 \rightarrow$ 4·84

$\bar{x} =$

$\therefore \quad F =$

Degrees of freedom are 10

$\therefore \quad F_{0\cdot05}$

$\therefore$ Variability of quality

Ex. 17-24. *Two random sa*
are characterized as follows :
Population from which the sample is drawn
I
II
You are to decide if the two

Sol. Let x and y be the obs

Now $\Sigma(x - \bar{x})^2 =$

$=$

and $\Sigma(y - \bar{y})^2 =$

$\therefore \quad S_1^2 =$

$\therefore \quad F =$

Nos. of *d.f.* are $8 - 1 = 7$ a

$\therefore \quad F_{0\cdot05}$

$\therefore$ Variances of two pop

1. Two independent sample
the variable (weight in ou
Sample I : 9
Sample II : 10
Test whether the estimate
$F_{0\cdot05}$ for 7 and 6 *d.f.* is 4·
[**Ans.** Not significant]

			4	5	3	3	= 45
			– 0·5	0·5	– 1·5	– 1·5	
			0·25	0·25	2·25	2·25	= 16·5
Process $B : y \rightarrow$	8	5	7	8	3	2	
$Y = (y - \bar{y}) \rightarrow$	2·2	– 0·8	1·2	2·2	– 2·8	– 3·8	
$Y^2 \rightarrow$	4·84	0·64	1·44	4·84	7·84	14·44	
			7	6	5	7	= 58
			1·2	0·2	– 0·8	1·2	
			1·44	0·04	0·64	1·44	= 37·6

$$\bar{x} = 4 \cdot 5, \bar{y} = 5 \cdot 8, S_1^2 = \frac{16 \cdot 5}{9} \text{ and } S_2^2 = \frac{37 \cdot 6}{9}$$

$$\therefore \quad F = \frac{37 \cdot 6}{16 \cdot 5} \simeq 2 \cdot 28$$

Degrees of freedom are 10 – 1 = 9 and 10 – 1 = 9.

$$\therefore \quad F_{0 \cdot 05} = 3 \cdot 18 > F_{cal}$$

$\therefore$ Variability of quality may be taken to be the same for two processes.

Ex. 17-24. *Two random samples of sizes* 8 *and* 11, *drawn from two normal populations, are characterized as follows :*

Population from which the sample is drawn	*Size of the sample*	*Sum of observations*	*Sum of squares of observations*
I	8	9·6	61·52
II	11	16·5	73·26

You are to decide if the two populations can be taken to have the same variance.

Sol. Let x and y be the observations for two samples.

Now $$\Sigma(x - \bar{x})^2 = \Sigma x^2 - N_1 \bar{x}^2 = \Sigma(x)^2 - \frac{1}{N_1} (\Sigma x)^2$$

$$= 61 \cdot 52 - \frac{1}{8} (9 \cdot 6)^2 = 50$$

and $$\Sigma(y - \bar{y})^2 = 73 \cdot 26 - \frac{1}{11} (16 \cdot 5)^2 = 48 \cdot 51$$

$$\therefore \quad S_1^2 = \frac{50}{7} \text{ and } S_2^2 = \frac{48 \cdot 51}{10}$$

$$\therefore \quad F = \frac{S_1^2}{S_2^2} = \frac{500}{339 \cdot 57} \simeq 1 \cdot 47$$

Nos. of *d.f. are* 8 – 1 = 7 and 11 – 1 = 10

$$\therefore \quad F_{0 \cdot 05} = 3 \cdot 14 > F_{cal}$$

$\therefore$ Variances of two populations may be same.

EXERCISES

1. Two independent samples of 8 and 7 items respectively had the following values of the variable (weight in ounces)

Sample I :	9	11	13	11	15	9	12	14
Sample II :	10	12	10	14	9	8	10	

Test whether the estimates of the population variances differ significantly. (Given that $F_{0 \cdot 05}$ for 7 and 6 *d.f.* is 4·21).

[**Ans.** Not significant]

2. In two groups of ten children each, increase in weight due to two different diets in the same period were in pounds.

8, 5, 7, 8, 3, 2, 7, 6, 5, 7,
3, 7, 5, 6, 5, 4, 4, 5, 3, 6

Test whether the variances differ significantly. (Given that $F_{0.05}$ for 9 and 9 *d.f.* is 3·18). **[Ans.** Not significant]

3. Two random samples drawn from two normal populations are :

Sample I : 20, 16, 26, 27, 23, 22, 18, 24, 25 and 19
Sample II : 27, 33, 42, 35, 32, 34, 38, 28, 41, 43, 30 and 37

Test whether the two populations have the same variance. **[Ans.** 2·14 Not significant]

4. The students of the same age of two different colleges were tested for variability of intelligence. The *I.Q's* of 10 students from one college showed a variance of 20 and those of an equal number from the other college had a variance of 15. Discuss whether there is any significant difference in vaiability. **[Ans.** Not significant]

5. In one sample of 8 observations the sum of the squares of the deviations of the sample values from the sample mean was 84·4 and in other sample of 10 observations it was 102·6. Test whether this difference is significant at 5% level, given that $F_{0.05}$ for 7 and 9 degrees of freedom is 3·29. **[Ans.** Not significant]

6. Two samples of sizes 9 and 8 give the sum of squares of deviations from their respective means equal to 160 inches squares and 91 inches squares respectively. Can they be regarded as drawn from the same normal population ? **[Ans.** 1·54, Not significant]

7. Two random samples gave the following results :

Size	Mean	S.D.
10	3·0	2·9
12	4·0	3·2

Test whether the samples come from the same normal population.

8. Two chemists A and B repeat a protein analysis 20 times. If X_i and Y_i are the values obtained by A and B respectively and if

$$\Sigma X_i = 196, \ \Sigma X_i^2 = 1928, \ \Sigma Y_i = 205 \text{ and } \Sigma Y_i^2 = 2105$$

Determine whether there is a significant difference in precision between the two sets of results, the precision being measured by the inverse of the variance. **[Ans.** Not significant]

9. The means of two random samples of sizes 9 and 7 respectively are 196·42 and 198·82. Their respective variances are 26·94 and 18·73. Can the samples be regarded as drawn from the same normal population. **[Ans.** No]

10. Two random samples gave the following results :

	Size	*Sample mean*	*Sum of squares of deviations from the sample mean*
Sample I :	10	15	90
Sample II :	12	14	108

Test whether the samples come from the same normal population.

11. Show how you would use Snedecor F-test to decide whether the following two samples have been drawn from the same normal population :

	Size	*mean*	*Sum of squares of deviations from mean*
Sample I :	9	68	36
Sample II :	10	69	42

17.5-2. Test for the Significance of an Observed Multiple Correlation Coefficient

Consider a random sample of size n from a $(k + 1)$ variate normal population. Let R be the multiple correlation co-efficient of a variate with k other variates. Hypothesis to be tested is

"The multiple correlation c
Assuming this hypothesis, t

$F =$

follows F-distribution with k, n

17.5.3. Test for the Significan

Let (x_i, y_{ij}), $(i = 1, 2 \ldots h, j$
population and let

$N =$

Let η be the correlation rati
The hypothesis to be tested
"Population correlation rati

$F =$

follows F-distribution with $h - 1$

17.5.4. Testing the Linearity o

Let η be the correlation rati
arranged in h arrays, from a biv
The test statistic for testing

$F =$

which follows F-distribution wi

17.6. Fisher's z-Distribution

Fisher's variate is defined b

$z =$

where F follows F distribution
$\therefore \quad F =$

Dist. of z is $dP =$

This distribution is called F

17.7. z-tests

Tests of significance based
z-tables provide critical val
Significance is tested by c
certain level of significance (5%
If $z_{0.05} > z_{cal}$ the ratio is insig
Similarly for 1% level.
Some z-tests are as below :

due to two different diets in the

en that $F_{0\cdot05}$ for 9 and 9 *d.f.* is
[**Ans.** Not significant]
ations are :
24, 25 and 19
28, 41, 43, 30 and 37
:e. [**Ans.** 2·14 Not significant]
es were tested for variability of
ge showed a variance of 20 and
variance of 15. Discuss whether
[**Ans.** Not significant]
s of the deviations of the sample
ample of 10 observations it was
% level, given that $F_{0\cdot05}$ for 7 and
[**Ans.** Not significant]
f deviations from their respective
juares respectively. Can they be
? [**Ans.** 1·54, Not significant]

S.D.
2·9
3·2

al population.
times. If X_i and Y_i are the values

$= 205$ and $\Sigma Y_i^2 = 2105$
n precision between the two sets
rse of the variance.
[**Ans.** Not significant]
spectively are 196·42 and 198·82.
the samples be regarded as drawn
[**Ans.** No]

Sum of squares of deviations from the sample mean
90
108

al population.
hether the following two samples
:

Sum of squares of deviations from mean
36
42

ple Correlation Coefficient

riate normal population. Let R be
other variates. Hypothesis to be

"The multiple correlation co-eff. in the population is zero"

Assuming this hypothesis, the statistic

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k}$$

follows F-distribution with k, $n-k-1$ *d.f.*

17.5.3. Test for the Significance of an Observed Sample Correlation Ratio

Let (x_i, y_{ij}), $(i = 1, 2 \ldots h, j = 1, \ldots\ldots n_i)$ be a random sample from a bivariate normal population and let

$$N = \sum_{i=1}^{h} n_i$$

Let η be the correlation ratio of y on x.

The hypothesis to be tested is :

"Population correlation ratio is zero" Assuming this hypothesis, the statistic

$$F = \frac{\eta^2}{1-\eta^2} \, \frac{N-h}{h-1}$$

follows F-distribution with $h-1$ and $N-h$. *d.f.*

17.5.4. Testing the Linearity of Regression

Let η be the correlation ratio and r the correlation coefficient for a sample of size of N arranged in h arrays, from a bivariate normal population.

The test statistic for testing the hypothesis of linearity of regression is

$$F = \frac{\eta^2 - r^2}{1-\eta^2} \cdot \frac{N-h}{h-2}$$

which follows F-distribution with $h-2$ and $N-h$ *d.fs.*

17.6. Fisher's z-Distribution

Fisher's variate is defined by

$$z = \frac{1}{2} \log_e F$$

where F follows F distribution with ν_1, ν_2 *d.fs.*

$$\therefore \quad F = e^{2z}$$

Dist. of z is
$$dP = \frac{2\nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{e^{\nu_1 z}}{(\nu_2 + \nu_1 e^{2z})^{\frac{\nu_1+\nu_2}{2}}} dz, \; -\infty < z < \infty$$

This distribution is called Fisher's z-distribution.

17.7. z-tests

Tests of significance based on z-distribution are called z-tests.

z-tables provide critical values of z for various values of ν_1, ν_2 at 5% or 1% levels.

Significance is tested by comparing the calculated value of z with tabulated value at certain level of significance (5% or 1%). Rules of decision are :

If $z_{0\cdot05} > z_{cal}$ the ratio is insignificant and if $z_{0\cdot05} < z_{cal}$, the ratio is significant at 5% level. Similarly for 1% level.

Some z-tests are as below :

17.7-1. Test for Equality of Variance

Consider two independent random samples $x_1, x_2, \ldots\ldots x_{n_1}$ and $y_1, y_2, \ldots\ldots y_{n_2}$ drawn from normal populations. The hypothesis to be tested is :

'The population variances are same'

Assuming this hypothesis, the statistic

$$z = \frac{1}{2}\log_e \frac{S_1^2}{S_2^2} = \frac{1}{2}\left\{(\log_e 10)\log_{10}\left(\frac{S_1^2}{S_2^2}\right)\right\}$$

$$= \frac{(2\cdot3026)}{2}\log_{10}\frac{S_1^2}{S_2^2} = 1\cdot1513\,\log_{10}\frac{S_1^2}{S_2^2}$$

where $\quad S_1^2 = \dfrac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2$ and $S_2^2 = \dfrac{1}{n_2-1}\sum_{j=1}^{n_2}(y_j-\bar{y})^2$

follows z-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ *d.f.*

Ex. 17-25. *Show how you would use student's t-test and Fisher's z-test to decide whether the two sets of observations*

17, 27, 18, 25, 27, 29, 27, 23, 17,

16, 16, 20, 16, 20, 17, 15, 21

indicate samples drawn from the same universe.

Sol. Let x and y be the variables for two samples respectively.

x :	17	27	18	25	27	29	27	23	17	Total
$X = x - 23$:	-6	4	-5	2	4	6	4	0	-6 =	3
X^2 :	36	16	25	4	16	36	16	0	36 =	185
y :	16	16	20	16	20	17	15	21		
$Y = y - 16$:	0	0	4	0	4	1	-1	5	=	13
Y^2 :	0	0	16	0	16	1	1	25	=	59

$$\bar{x} = 23 + \frac{3}{9} = \frac{70}{3},\ \bar{y} = 16 + \frac{13}{8} = \frac{141}{8}$$

$$\Sigma(x-\bar{x})^2 = \Sigma(X-\bar{X})^2 = \Sigma X^2 - \frac{1}{n_1}(\Sigma X)^2$$

$$= 185 - \frac{1}{9}(9) = 184$$

and

$$\Sigma(y-\bar{y})^2 = \Sigma(Y-\bar{Y})^2$$

$$= (\Sigma Y^2) - \frac{1}{n_2}(\Sigma Y)^2 = 59 - \frac{1}{8}(169) = \frac{303}{8}$$

$$\therefore \quad S_1^2 = \frac{184}{8},\ S_2^2 = \frac{303}{56}.$$

Firstly the equality of population variances will be tested by applying z-test.

Now $$z = 1\cdot1513\,\log_{10}\frac{S_1^2}{S_2^2} = (1\cdot1513)\log_{10}\frac{1288}{303}$$

$$= (1\cdot1513)\{\log_{10}1288 - \log_{10}303\}$$

$$= (1\cdot1513)\{3\cdot1099 - 2\cdot4814\} \simeq 0\cdot724$$

Now $z_{0\cdot05}$ for 8 and 7 *d.f.* $= 0\cdot6576 < z_{\text{cal}}$

and $z_{0\cdot01}$ for 8 and 7 *d.f.* =

$\therefore$ At 5% level the varia

$\therefore$ At 1% level, the two

Now t-test will be applied t

S^2 :

$\therefore$ $|t|$ =

No. of *d.f.* =

$\therefore$ $t_{0\cdot05}$ =

$\therefore$ t_{cal} >

$\therefore$ The difference betwe

$\therefore$ The two samples do n

Ex. 17-26. (*i*) *Give the test z-transformation.*

(*ii*) *A correlation co-efficie transformation to find out if this*

Sol. (*i*) Let 'r' and 'ρ' be respectively and n the sample si

Fisher z-transformation is

z =

For large values of 'n', 'z' where

ξ =

and variance $\dfrac{1}{n-3}$

$\therefore$ For large values of n,

u =

is asymptotically standard norma by calculating 'u' and using nor

Thus if $|u| > 1\cdot96$, the differe is significant at 1% and if $|u| > 3$

Note. The symbol 'z' used

and $z_{0.01}$ for 8 and 7 $d.f. = 0.9614 > z_{cal}$

$\therefore$ At 5% level the variance ratio is significant and at 1% level not significant.

$\therefore$ At 1% level, the two population variances may be taken to be same.

Now t-test will be applied to test the significance of the difference between the means.

$$S^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{j=1}^{n_2} (y_j - \bar{y})^2}{n_1 + n_2 - 2} = \frac{184 + \frac{303}{8}}{15}$$

$$= \frac{1775}{120}$$

$$\therefore \qquad |t| = \frac{\bar{x} \sim \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\frac{70}{3} - \frac{141}{8}}{\sqrt{\frac{1775}{120}}\sqrt{\frac{1}{8} + \frac{1}{9}}} \sim 3.05$$

$$\text{No. of } d.f. = 9 + 8 - 2 = 15$$

$$\therefore \qquad t_{0.05} = 2.13 \text{ and } t_{0.01} = 2.95$$

$$\therefore \qquad t_{cal} > t_{0.01} \text{ and } t_{0.05}$$

$\therefore$ The difference between the means is significant.

$\therefore$ The two samples do not belong to the same universe.

Ex. 17·26. (*i*) *Give the test of significance of correlation co-efficient based on Fisher z-transformation.*

(*ii*) *A correlation co-efficient of* 0·7 *is discovered in a sample of* 28 *pairs. Apply z-transformation to find out if this differs significantly* (*a*) *from* 0; (*b*) *from* 0·5.

Sol. (*i*) Let 'r' and 'ρ' be the correlation co-efficient for sample and the population respectively and n the sample size

Fisher z-transformation is

$$z = \frac{1}{2}\log_e \frac{1+r}{1-r} = 1.1513 \log_{10} \frac{1+r}{1-r}$$

For large values of 'n', 'z' is distributed asymptotically normally about the mean 'ξ' where

$$\xi = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} = 1.1513 \log_{10} \frac{1+\rho}{1-\rho}$$

and variance $\dfrac{1}{n-3}$

$\therefore$ For large values of n, the variate

$$u = \frac{z - \xi}{\sqrt{\frac{1}{n-3}}}$$

is asymptotically standard normal variate. Thus the significance between 'r' and 'ρ' is tested by calculating 'u' and using normal tables.

Thus if $|u| > 1.96$, the difference is significant at 5% level and if $|u| > 2.58$, the difference is significant at 1% and if $|u| > 3$, the difference is highly significant.

Note. The symbol 'z' used here is different from Fisher's z-distribution.

(*ii*) (*a*)

$$n = 28, \quad \rho = 0, \quad r = 0{\cdot}7$$

$\therefore$

$$\xi = 0, \quad z = (1{\cdot}1513)\log_{10}\frac{1{\cdot}7}{0{\cdot}3}$$

$$= (1{\cdot}1513)\{\log_{10}17 - \log_{10}3\}$$

$$= (1{\cdot}1513)\{1{\cdot}2304 - 0{\cdot}4771\} = 0{\cdot}87$$

$\therefore$ $$u = (0{\cdot}87)\sqrt{25} = 4{\cdot}35 > 3.$$

$\therefore$ The hypothesis of zero correlation is refuted and hence the population is correlated.

(*b*)

$$n = 28, \quad \rho = 0{\cdot}5, \quad r = 0{\cdot}7$$

$\therefore$

$$\xi = (1{\cdot}1513)\log_{10}\frac{1{\cdot}5}{0{\cdot}5} = (1{\cdot}1513)\log_{10}3$$

$$= (1{\cdot}1513)(0{\cdot}4771) \simeq 0{\cdot}55$$

Also $$z = 0{\cdot}87$$

$\therefore$ $$u = (0{\cdot}32)\sqrt{25} = 1{\cdot}60 < 1{\cdot}96.$$

$\therefore$ The difference between r and ρ is not significant and hence the hypothesis that $\rho = 0{\cdot}5$ is acceptable.

Ex. 17-27. *What is the probability that a correlation co-efficient of* $0{\cdot}75$ *or less can arise in a sample of* 30 *from a normal population in which the true correlation is* $0{\cdot}9$?

Sol. $$r = 0{\cdot}75, \quad \rho = 0{\cdot}9, \quad n = 30$$

$\therefore$

$$\xi = (1{\cdot}1513)\log_{10}\frac{1{\cdot}9}{0{\cdot}1} = 1{\cdot}1513\log_{10}19$$

$$= (1{\cdot}1513)(1{\cdot}2788) \simeq 1{\cdot}472$$

$$z = (1{\cdot}1513)\log_{10}\left(\frac{1{\cdot}75}{0{\cdot}25}\right) = (1{\cdot}1513)\log_{10}7$$

$$= (1{\cdot}1513)(0{\cdot}8451) \simeq 0{\cdot}973$$

$\therefore$ $$u = (0{\cdot}973 - 1{\cdot}472)\sqrt{27} \simeq -2{\cdot}59$$

Now $$P\{r \le 0{\cdot}75\} = P\{1 + r \le 1{\cdot}75\} = P\left\{\frac{1+r}{1-r} \le \frac{1{\cdot}75}{0{\cdot}25} = 7\right\}$$

$$= P\{z \le 0{\cdot}973\} = P\{u < -2{\cdot}59\}$$

$$= 0{\cdot}5 - P\{-2{\cdot}59 < u < 0\} = 0{\cdot}5 - P\{0 < u < 2{\cdot}59\}$$

$$= 0{\cdot}5 - 0{\cdot}4952 = 0{\cdot}0048.$$

Ex. 17-28. *In a random sample of* 19 *pairs of values from a bivariate normal population, the correlation was found to be* $0{\cdot}7$. *Is this value consistent with the assumption that the correlation in the population is* $0{\cdot}5$? [**Ans.** $u = 1{\cdot}28$]

Ex. 17-29. (*a*) *Give the procedure of testing the significance of the difference between two independent correlation co-efficients.*

(*b*) *The correlation coefficient between temperature of rice and breakage percentage calculated from two samples of* 12 *and* 16 *are* $0{\cdot}8912$ *and* $0{\cdot}8482$ *respectively. Do the two estimates differ significantly* ?

Sol. (*a*) Let there be two independent random samples of sizes n_1, n_2 and correlation co-efficients r_1, r_2. The hypothesis to be tested is :

'Can the samples be regarded as drawn from the same population or from two populations with the same correlation co-efficients'.

Assuming this hypothesis, the statistic

$$u = \frac{z_1 - z_2}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$$

where $z_i = (1$

is asymptotically standard norma

Thus the significance of the (

using normal tables.

(*c*) $n_1 = 1$

$z_1 = (1$

$= (1$

$= (1$

$\simeq 1\cdot$

$z_2 = (1\cdot$

$= (1\cdot$

$= (1\cdot$

$u =$

$\therefore$ Difference between r_1 a

Ex. 17-30. *The first of two sa*

while the second of 28 *pairs has a c*

Sol. $n_1 = 23,$

$\therefore$ $z_1 = (1\cdot1$

$= (1\cdot1$

$z_2 = (1\cdot1$

$= (1\cdot1$

$|u| = 1\cdot0$

$\therefore$ Difference is not signific

Ex. 17-31. *The correlation co*

aptitude for a group of 20 *girls is* $0\cdot$

significant ?

where $$z_i = (1 \cdot 1513) \log_{10} \frac{1+r_i}{1-r_i} \qquad (i = 1, 2)$$
is asymptotically standard normal variate.

Thus the significance of the difference between r_1 and r_2 is tested by calculating u and using normal tables.

(*c*) $$n_1 = 12,\ n_2 = 16,\ r_1 = 0 \cdot 8912,\ r_2 = 0 \cdot 8482$$

$$\begin{aligned} z_1 &= (1 \cdot 1513) \left\{ \log_{10} \frac{1 \cdot 8912}{0 \cdot 1088} \right\} \\ &= (1 \cdot 1513) \{ \log_{10} 18912 - \log_{10} 1088 \} \\ &= (1 \cdot 1513) \{ 4 \cdot 2767 - 3 \cdot 0366 \} = (1 \cdot 1513)(1 \cdot 2401) \\ &\simeq 1 \cdot 428 \end{aligned}$$

$$\begin{aligned} z_2 &= (1 \cdot 1513) \left\{ \log_{10} \frac{1 \cdot 8482}{0 \cdot 1518} \right\} \\ &= (1 \cdot 1513) \{ \log_{10} 18482 - \log_{10} 1518 \} \\ &= (1 \cdot 1513) \{ 4 \cdot 2667 - 3 \cdot 1813 \} = (1 \cdot 1513)(1 \cdot 0854) \simeq 1 \cdot 250. \end{aligned}$$

$$u = \frac{1 \cdot 428 - 1 \cdot 250}{\sqrt{\dfrac{1}{12-3} + \dfrac{1}{16-3}}} \simeq 0 \cdot 41 < 1 \cdot 96.$$

$\therefore$ Difference between r_1 and r_2 is not significant.

Ex. 17-30. *The first of two samples consists of 23 pairs and gives a correlation of* $0 \cdot 5$ *while the second of 28 pairs has a correlation of* $0 \cdot 8$. *Are these values significantly different ?*

Sol. $$n_1 = 23,\ n_2 = 28,\quad r_1 = 0 \cdot 5 \quad \text{and} \quad r_2 = 0 \cdot 8$$

$$\begin{aligned} \therefore \quad z_1 &= (1 \cdot 1513) \left\{ \log_{10} \frac{1 \cdot 5}{0 \cdot 5} \right\} = (1 \cdot 1513) \log_{10} 3 \\ &= (1 \cdot 1513)(0 \cdot 4771) \simeq 0 \cdot 5493 \\ z_2 &= (1 \cdot 1513) \log_{10} \frac{1 \cdot 8}{0 \cdot 2} = 1 \cdot 1513 \log_{10} 9 \\ &= (1 \cdot 1513)(0 \cdot 9542) \simeq 1 \cdot 0986 \end{aligned}$$

$$|u| = \frac{1 \cdot 0986 - 0 \cdot 5493}{\sqrt{\dfrac{1}{20} + \dfrac{1}{25}}} \simeq 1 \cdot 83 < 1 \cdot 96$$

$\therefore$ Difference is not significant.

Ex. 17-31. *The correlation co-efficient between Mathematics aptitude and Physics aptitude for a group of 20 girls is* $0 \cdot 42$ *and for a group of 25 boys is* $0 \cdot 75$. *Is the difference significant ?*

[Ans. $1 \cdot 6$, Not significant**]**

	0	1	2	3	4	
10	0000	0043	0086	0128	0170	
11	0414	0453	0492	0531	0569	
12	0792	0828	0864	0899	0934	
13	1139	1173	1206	1239	1271	
14	1461	1492	1523	1553	1584	
15	1761	1790	1818	1847	1875	1
16	2041	2068	2095	2122	2148	2
17	2304	2330	2355	2380	2405	2
18	2553	2577	2601	2625	2648	2
19	2788	2810	2833	2856	2878	2
20	3010	3032	3054	3075	3096	3
21	3222	3243	3263	3284	3304	3
22	3424	3444	3464	3483	3502	3
23	3617	3636	3655	3674	3692	3
24	3902	3820	3838	3856	3874	3
25	3979	3997	4014	4031	4048	4
26	4150	4166	4183	4200	4216	4
27	4341	4330	4346	4362	4378	4
28	4472	4487	4502	4518	4533	4
29	4624	4639	4654	4669	4683	4
30	4771	4786	4800	4814	4829	48
31	4914	4928	4942	4955	4969	49
32	5051	5065	5079	5092	5105	51
33	5185	5198	5211	5224	5237	52
34	5315	5328	5340	5353	5366	53
35	5441	5433	5465	5478	5490	55
36	5563	5575	5587	5599	5611	56
37	5682	5694	5705	5717	5729	57
38	5798	5809	5821	5832	5843	58
39	5911	5922	5933	5944	5955	59
40	6021	6031	6042	6053	6064	60
41	6128	6138	6149	6160	6170	61
42	6232	6243	6253	6263	6274	62
43	6335	6345	6355	6365	6375	63
44	6435	6444	6454	6464	6474	64
45	6332	6542	6551	6561	6571	65
46	6628	6637	6646	6656	6665	66
47	6721	6730	6739	6749	6758	67
48	6812	6821	6830	6839	6848	68
49	6902	6911	6920	6928	6937	69

Table 1

LOGARITHMS

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170						5	9	13	17	21	26	30	34	38
						0212	0253	0294	0334	0374	4	8	12	16	20	24	28	32	36
11	0414	0453	0492	0531	0569						4	8	12	16	20	23	27	31	35
						0607	0645	0682	0719	0755	4	7	11	15	18	22	26	29	33
12	0792	0828	0864	0899	0934						3	7	11	14	18	21	25	28	32
						0969	1004	1038	1072	1106	3	7	10	14	17	20	24	27	31
13	1139	1173	1206	1239	1271						3	6	10	13	16	19	23	26	29
						1303	1335	1367	1399	1430	3	7	10	13	16	19	22	25	29
14	1461	1492	1523	1553	1584						3	6	9	12	15	19	22	25	29
						1614	1644	1673	1703	1732	3	6	9	12	14	17	20	23	26
15	1761	1790	1818	1847	1875						3	6	9	11	14	17	20	23	26
						1903	1931	1959	1987	2014	3	6	8	11	14	17	19	22	25
16	2041	2068	2095	2122	2148						3	6	8	11	14	16	19	22	24
						2175	2201	2227	2253	2279	3	5	8	10	13	16	18	21	23
17	2304	2330	2355	2380	2405						3	5	8	10	13	15	18	20	23
						2430	2455	2480	2504	2329	3	5	8	10	12	15	17	20	22
18	2553	2577	2601	2625	2648						2	5	7	9	12	14	17	19	21
						2672	2695	2718	2742	2765	2	4	7	9	11	14	16	18	21
19	2788	2810	2833	2856	2878						2	4	7	9	11	13	14	18	20
						2900	2923	2945	2967	2989	2	4	6	8	11	13	15	17	19
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3902	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4341	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	8	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5433	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	9	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6332	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8

Table II

LOGARITHMS

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	3	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7784	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	6	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	6
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

	0	1	2	3	4
·00	1000	1002	1005	1007	1009
·01	1023	1026	1028	1030	1033
·02	1047	1050	1052	1054	1057
·03	1072	1074	1076	1079	1081
·04	1096	1099	1102	1104	1107
·05	1122	1125	1127	1130	1132
·06	1148	1151	1153	1156	1159
·07	1175	1178	1180	1183	1186
·08	1202	1205	1208	1211	1213
·09	1230	1233	1236	1239	1242
·10	1259	1262	1265	1268	1271
·11	1288	1291	1294	1297	1300
·12	1318	1321	1324	1327	1330
·13	1349	1352	1355	1358	1361
·14	1380	1384	1387	1390	1393
·15	1413	1416	1419	1422	1426
·16	1445	1449	1452	1455	1459
·17	1479	1483	1486	1489	1493
·18	1514	1517	1521	1524	1528
·19	1549	1552	1556	1560	1563
·20	1585	1589	1592	1596	1600
·21	1622	1626	1629	1633	1637
·22	1660	1663	1667	1671	1675
·23	1698	1702	1706	1710	1714
·24	1738	1742	1746	1750	1754
·25	1778	1782	1786	1791	1795
·26	1820	1824	1828	1832	1837
·27	1862	1866	1871	1875	1879
·28	1905	1910	1914	1919	1923
·29	1950	1954	1959	1963	1968
·30	1995	2000	2004	2009	2014
·31	2042	2046	2051	2056	2061
·32	2089	2094	2099	2104	2109
·33	2138	2143	2148	2153	2158
·34	2188	2193	2198	2203	2208
·35	2239	2244	2249	2254	2259
·36	2291	2296	2301	2307	2312
·37	2344	2350	2355	2360	2366
·38	2399	2404	2410	2415	2421
·39	2455	2460	2466	2472	2477
·40	2512	2518	2523	2529	2535
·41	2570	2576	2582	2588	2594
·42	2630	2636	2642	2649	2655
·43	2692	2698	2704	2710	2716
·44	2754	2761	2767	2773	2780
·45	2818	2825	2831	2838	2844
·46	2884	2891	2897	2904	2911
·47	2951	2958	2965	2972	2979
·48	3020	3027	3034	3041	3048
·49	3090	3097	3105	3112	3119

9	1	2	3	4	5	6	7	8	9
067	1	2	3	3	4	5	6	7	8
152	1	2	3	3	4	5	6	7	8
235	1	2	3	3	4	5	6	7	7
316	1	2	2	3	4	5	6	6	7
396	1	2	2	3	4	5	6	6	7
474	1	2	2	3	4	5	5	6	7
551	1	2	2	3	4	5	5	6	7
627	1	2	2	3	4	5	5	6	7
701	1	1	2	3	4	4	5	6	7
774	1	1	2	3	4	6	5	6	7
846	1	1	2	3	4	4	5	6	6
917	1	1	2	3	4	4	5	6	6
987	1	1	2	3	3	4	5	6	6
055	1	1	2	3	3	4	5	5	6
122	1	1	2	3	3	4	5	5	6
189	1	1	2	3	3	4	5	5	6
254	1	1	2	3	3	4	5	5	6
319	1	1	2	3	3	4	5	5	6
382	1	1	2	3	3	4	4	5	6
445	1	1	2	2	3	4	4	5	6
506	1	1	2	2	3	4	4	5	6
567	1	1	2	2	3	4	4	5	5
627	1	1	2	2	3	4	4	5	5
686	1	1	2	2	3	4	4	5	5
745	1	1	2	2	3	4	4	5	6
802	1	1	2	2	3	3	4	5	5
859	1	1	2	2	3	3	4	5	5
915	1	1	2	2	3	3	4	4	5
971	1	1	2	2	3	3	4	4	5
025	1	1	2	2	3	3	4	4	5
079	1	1	2	2	3	3	4	4	5
133	1	1	2	2	3	3	4	4	5
186	1	1	2	2	3	3	4	4	5
238	1	1	2	2	3	3	4	4	5
289	1	1	2	2	3	3	4	4	5
340	1	1	2	2	3	3	4	4	5
390	1	1	2	2	3	3	4	4	5
440	0	1	1	2	2	3	3	4	4
489	0	1	1	2	2	3	3	4	4
538	0	1	1	2	2	3	3	4	4
9586	0	1	1	2	2	3	3	4	4
9633	0	1	1	2	2	3	3	4	4
9680	0	1	1	2	2	3	3	4	4
9727	0	1	1	2	2	3	3	4	4
9773	0	1	1	2	2	3	3	4	4
9818	0	1	1	2	2	3	3	4	4
9863	0	1	1	2	2	3	3	4	4
9908	0	1	1	2	2	3	3	4	4
9952	0	1	1	2	2	3	3	4	4
9996	0	1	1	2	2	3	3	3	4

Table 1II

ANTILOGARITHMS

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
·11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
·14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	2	3	3
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2285	1	1	2	2	3	3	4	4	5
·36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
·37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
·39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
·40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
·43	2692	2698	2704	2710	2716	2723	2629	2735	2742	2748	1	1	2	3	3	4	4	5	6
·44	2754	2761	2767	2773	2780	2784	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	3	4	5	6	6
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	3	4	5	6	6

Table 1V

ANTILOGARITHMS

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
·57	3715	3724	3733	3741	3750	3758	3767	3776	3484	3793	1	2	3	3	4	5	6	7	8
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
·66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
·79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
·98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
·99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

HYPERB

	0	1	2	3	
1·0	0·0000	0099	0198	0296	03
1·1	·0953	1044	1133	1222	13
1·2	·1823	1906	1989	2070	21
1·3	·2624	2700	2776	2852	29
1·4	·3365	3436	3507	3577	36
1·5	·4055	4121	4187	4253	43
1·6	·4700	4762	4824	4886	49
1·7	·5306	5365	5423	5481	55
1·8	·5878	5933	5988	6043	60
1·9	·6419	6471	6523	6575	66
2·0	·6931	6981	7031	7080	71
2·1	·7419	7467	7514	7561	76
2·2	·7885	7930	7975	8020	80
2·3	·8329	8372	8416	8459	85
2·4	·8755	8796	8838	8879	89
2·5	·9163	9203	9243	9282	93
2·6	·9555	9594	9632	9670	97
2·7	·9933	9969	1.0006	0043	00
2·8	1·0296	0332	0367	0403	04
2·9	1·0647	0682	0716	0750	07
3·0	1·0986	1019	1053	1086	11
3·1	1·1314	1346	1378	1410	14
3·2	1·1632	1663	1694	1725	17
3·3	1·1939	1969	1.2000	2030	20
3·4	1·2238	2267	2296	2326	23
3·5	1·2528	2556	2585	2613	26
3·6	1·2809	2837	2865	2892	29
3·7	1·3083	3110	3137	3164	31
3·8	1·3350	3376	3403	3429	34
3·9	1·3610	3635	3661	3686	37
4·0	1·3863	3888	3913	3938	39
4·1	1·4110	4134	4159	4183	42
4·2	1·4351	4375	4398	4422	44
4·3	1·4586	4609	5633	4656	46
4·4	1·4816	4839	4861	4884	49
4·5	1·5041	5063	5085	5107	51
4·6	1·5261	5282	5304	5326	53
4·7	1·5476	5497	5518	5539	55
4·8	1·5686	5707	5728	5748	57
4·9	1·5892	5913	5933	5953	59
5·0	1·6094	6114	6134	6154	61
5·1	1·6292	6312	6332	6351	63
5·2	1·6487	6506	6525	6544	65
5·3	1·6677	6696	6715	6734	67
5·4	1·6864	6882	6901	6919	69

Hyperb

n	1	2	
$\log_e 10^n$	2·3026	4·6052	6

S

9	1	2	3	4	5	6	7	8	9
3228	1	1	2	3	4	4	5	6	7
3304	1	2	2	3	4	5	5	6	7
3381	1	2	2	3	4	5	5	6	7
3459	1	2	2	3	4	5	6	6	7
3540	1	2	2	3	4	5	6	6	7
3622	1	2	2	3	4	5	6	7	7
3707	1	2	3	3	4	5	6	7	8
3793	1	2	3	3	4	5	6	7	8
3882	1	2	3	4	4	5	6	7	8
3972	1	2	3	4	5	5	6	7	8
4064	1	2	3	4	5	6	6	7	8
4159	1	2	3	4	5	6	7	8	9
4256	1	2	3	4	5	6	7	8	9
4355	1	2	3	4	5	6	7	8	9
4457	1	2	3	4	5	6	7	8	9
4560	1	2	3	4	5	6	7	8	9
4667	1	2	3	4	5	6	7	9	10
4775	1	2	3	4	5	7	8	9	10
4887	1	2	3	4	6	7	8	9	10
5000	1	2	3	5	6	7	8	9	10
5117	1	2	4	5	6	7	8	9	11
5236	1	2	4	5	6	7	8	10	11
5358	1	2	4	5	6	7	9	10	11
5483	1	3	4	5	6	8	9	10	11
5610	1	3	4	5	6	8	9	10	12
5741	1	3	4	5	7	8	9	10	12
5875	1	3	4	5	7	8	9	11	12
6012	1	3	4	5	7	8	10	11	12
6152	1	3	4	6	7	8	10	11	13
6295	1	3	4	6	7	9	10	11	13
6442	1	3	4	6	7	9	10	12	13
6592	2	3	5	6	8	9	11	12	14
6745	2	3	5	6	8	9	11	12	14
6902	2	3	5	6	8	9	11	13	14
7063	2	3	5	6	8	10	11	13	15
7228	2	3	5	7	8	10	12	13	15
7396	2	3	5	7	8	10	12	13	15
7568	2	3	5	7	9	10	12	14	16
7745	2	4	5	7	9	11	12	14	16
7925	2	4	5	7	9	11	13	14	16
8110	2	4	6	7	9	11	13	15	17
8299	2	4	6	8	9	11	13	15	17
8492	2	4	6	8	10	12	14	15	17
8690	2	4	6	8	10	12	14	16	18
8892	2	4	6	8	10	12	14	16	18
9099	2	4	6	8	10	12	15	17	19
9311	2	4	6	8	11	13	15	17	19
9528	2	4	7	9	11	13	15	17	20
9750	2	4	7	9	11	13	16	18	20
9977	2	5	7	9	11	14	16	18	20

Table V

HYPERBOLIC OR NAPERIAN LOGARITHMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
1·0	0·0000	0099	0198	0296	0392	0488	0583	0677	0770	0862	10	19	29	38	48	57	67	76	86
1·1	·0953	1044	1133	1222	1310	1398	1484	1570	1655	1740	9	17	26	35	44	52	61	70	78
1·2	·1823	1906	1989	2070	2151	2231	2311	2390	2469	2546	8	16	24	32	40	48	56	64	72
1·3	·2624	2700	2776	2852	2927	3001	3075	3148	3221	3293	7	15	22	30	37	44	52	59	67
1·4	·3365	3436	3507	3577	3646	3716	3784	3853	3920	3988	7	14	21	28	35	41	48	55	62
1·5	·4055	4121	4187	4253	4318	4383	4447	4511	4574	4637	6	13	19	26	32	39	45	52	58
1·6	·4700	4762	4824	4886	4947	5008	5068	5128	5188	5247	6	12	18	24	30	36	42	48	55
1·7	·5306	5365	5423	5481	5539	5596	5653	5710	5766	5822	6	11	17	24	29	34	40	46	51
1·8	·5878	5933	5988	6043	6098	6152	6206	6259	6313	6366	5	11	16	22	27	32	38	43	49
1·9	·6419	6471	6523	6575	6627	6678	6729	6780	6831	6881	5	10	15	20	26	31	36	41	46
2·0	·6931	6981	7031	7080	7129	7178	7227	7275	7324	7372	5	10	15	20	24	29	34	39	44
2·1	·7419	7467	7514	7561	7608	7655	7701	7747	7793	7839	5	9	14	19	23	28	33	37	42
2·2	·7885	7930	7975	8020	8065	8109	8154	8198	8242	8286	4	9	13	18	22	27	31	36	40
2·3	·8329	8372	8416	8459	8502	8544	8587	8629	8671	8713	4	9	13	17	21	26	30	34	38
2·4	·8755	8796	8838	8879	8920	8961	9002	9042	9083	9123	4	8	12	16	20	24	29	33	37
2·5	·9163	9203	9243	9282	9322	9361	9400	9439	9478	9517	4	8	12	16	20	24	27	31	35
2·6	·9555	9594	9632	9670	9708	9746	9783	9821	9858	9895	4	8	11	15	19	23	26	30	34
2·7	·9933	9969	1.0006	0043	0080	0116	0152	0188	0225	0260	4	7	11	15	18	22	25	29	33
2·8	1·0296	0332	0367	0403	0438	0473	0508	0543	0578	0613	4	7	11	14	18	21	25	28	32
2·9	1·0647	0682	0716	0750	0784	0818	0852	0886	0919	0953	3	7	10	14	17	20	24	27	31
3·0	1·0986	1019	1053	1086	1119	1151	1184	1217	1249	1282	3	7	10	13	16	20	23	26	30
3·1	1·1314	1346	1378	1410	1442	1474	1506	1537	1569	1600	3	6	10	13	16	19	22	25	29
3·2	1·1632	1663	1694	1725	1756	1787	1817	1848	1878	1909	3	6	9	12	15	18	22	25	28
3·3	1·1939	1969	1.2000	2030	2060	2090	2119	2149	2179	2208	3	6	9	12	15	18	21	24	27
3·4	1·2238	2267	2296	2326	2355	2384	2413	2442	2470	2499	3	6	9	12	15	17	20	23	26
3·5	1·2528	2556	2585	2613	2641	2669	2698	2726	2754	2682	3	6	8	11	14	17	20	23	25
3·6	1·2809	2837	2865	2892	2920	2947	2975	3002	3029	3056	3	5	8	11	14	16	19	22	25
3·7	1·3083	3110	3137	3164	3191	3218	3244	3271	3297	3324	3	5	8	11	13	16	19	21	24
3·8	1·3350	3376	3403	3429	3455	3481	3507	3533	3558	3584	3	5	8	10	13	16	18	21	23
3·9	1·3610	3635	3661	3686	3712	3737	3762	3788	3813	3838	3	5	8	10	13	15	18	20	23
4·0	1·3863	3888	3913	3938	3962	3987	4012	4036	4061	4085	2	5	7	10	12	15	17	20	22
4·1	1·4110	4134	4159	4183	4207	4231	4255	4279	4303	4327	2	5	7	10	12	14	17	19	22
4·2	1·4351	4375	4398	4422	4446	4469	4493	4516	4540	4563	2	5	7	9	12	14	16	19	21
4·3	1·4586	4609	5633	4656	4679	4702	4725	4748	4770	4793	2	5	7	9	12	14	16	18	21
4·4	1·4816	4839	4861	4884	4907	4929	4951	4974	4996	5019	2	5	7	9	11	14	16	18	20
4·5	1·5041	5063	5085	5107	5129	5151	5173	5195	5217	5239	2	4	7	9	11	13	15	18	20
4·6	1·5261	5282	5304	5326	5347	5369	5390	5412	5433	5454	2	4	6	9	11	13	15	17	19
4·7	1·5476	5497	5518	5539	5560	5581	5602	5623	5644	5665	2	4	6	8	11	13	15	17	19
4·8	1·5686	5707	5728	5748	5769	5790	5810	5831	5851	5872	2	4	6	8	10	12	14	16	19
4·9	1·5892	5913	5933	5953	5974	5994	6014	6034	6054	6074	2	4	6	8	10	12	14	16	18
5·0	1·6094	6114	6134	6154	6174	6194	6214	6233	6253	6273	2	4	6	8	10	12	14	16	18
5·1	1·6292	6312	6332	6351	6371	6390	6409	6429	6448	6467	2	4	6	8	10	12	14	16	18
5·2	1·6487	6506	6525	6544	6563	6582	6601	6620	6639	6658	2	4	6	8	10	11	13	15	17
5·3	1·6677	6696	6715	6734	6752	6771	6790	6808	6827	6845	2	4	6	7	9	11	13	15	17
5·4	1·6864	6882	6901	6919	6938	6956	6974	6993	7011	7029	2	4	5	7	9	11	13	15	17

Hyperbolic or Naperian Logarithms of 10^{+n}.

n	1	2	3	4	5	6	7	8	9
$\log_e 10^n$	2·3026	4·6052	6·9078	9·2103	11·5129	13·8155	16·1181	18·4207	20·7233

Table VI

HYPERBOLIC OR NAPERIAN LOGARITHMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
5·5	1·7047	7066	7084	7102	7120	7138	7156	7174	7192	7210	2	4	5	7	9	11	13	14	16
5·6	1·7228	7246	7263	7381	7299	7317	7334	7352	7370	7387	2	4	5	7	9	11	12	14	16
5·7	1·7405	7422	7440	7457	7475	7492	7509	7527	7544	7561	2	3	5	7	9	10	12	14	16
5·8	1·7579	7596	7613	7630	7647	7664	7681	7699	7716	7733	2	3	5	7	9	10	12	14	15
5·9	1·7750	7766	7783	7800	7817	7834	7851	7867	7884	7901	2	3	5	7	8	10	12	13	15
6·0	1·7918	7934	7951	7967	7984	8001	8017	8034	8050	8066	2	3	5	7	8	10	12	13	15
6·1	1·8083	8099	8116	8132	8148	8165	8181	8197	8213	8229	2	3	5	6	8	10	11	13	15
6·2	1·8245	8262	8278	8294	8310	8326	8342	8358	8374	8390	2	3	5	6	8	10	11	13	14
6·3	1·8405	8421	8437	8453	8469	8485	8500	8516	8532	8547	2	3	5	6	8	9	11	13	14
6·4	1·8563	8579	8594	8610	8625	8641	8656	8672	8687	8703	2	3	5	6	8	9	11	12	14
6·5	1·8718	8733	8749	8764	8779	8795	8810	8825	8840	8856	2	3	5	6	8	9	11	12	14
6·6	1·8871	8886	8901	8916	8931	8946	8961	8976	8991	9006	2	3	5	6	8	9	11	12	14
6·7	1·9021	9036	9051	9066	9081	9095	9110	9125	9140	9155	1	3	4	6	7	9	10	12	13
6·8	1·9169	9184	9199	9213	9228	9242	9257	9272	9286	9301	1	3	4	6	7	9	10	12	13
6·9	1·9315	9330	9344	9359	9373	9387	9402	9416	9430	9445	1	3	4	6	7	9	10	12	13
7·0	1·9459	9473	9488	9502	9516	9530	9544	9559	9573	9587	1	3	4	6	7	9	10	11	13
7·1	1·9601	9615	9629	9643	9657	9671	9685	9699	9713	9727	1	3	4	6	7	8	10	11	13
7·2	1·9741	9755	9769	9782	9796	9810	9824	9868	9851	9865	1	3	4	6	7	8	10	11	12
7·3	1·9879	9892	9906	9920	9933	9947	9961	9974	9988	2.0001	1	3	4	5	7	8	10	11	12
7·4	2·0015	0028	0042	0055	0069	0082	0096	0109	0122	0136	1	3	4	5	7	8	9	11	12
7·5	2·0149	0162	0176	0189	0202	0215	0229	0242	0255	0268	1	3	4	5	7	8	9	11	12
7·6	2·0281	0295	0308	0321	0334	0347	0360	0373	0386	0399	1	3	4	5	7	8	9	10	12
7·7	2·0412	0425	0438	0451	0464	0477	0490	0503	0516	0528	1	3	4	5	6	8	9	10	12
7·8	2·0541	0554	0567	0580	0592	0605	0618	0631	0643	0656	1	3	4	5	6	8	9	10	11
7·9	2·0669	0681	0694	0707	0719	0732	0744	0757	0769	0782	1	3	4	5	6	8	9	10	11
8·0	2·0794	0807	0819	0832	0844	0857	0869	0882	0894	0906	1	3	4	5	6	7	9	10	11
8·1	2·0919	0931	0943	0956	0968	0980	0992	1005	1017	1029	1	2	4	5	6	7	9	10	11
8·2	2·1041	1054	1066	1078	1090	1102	1114	1126	1138	1150	1	2	4	5	6	7	9	10	11
8·3	2·1163	1175	1187	1199	1211	1223	1235	1247	1258	1270	1	2	4	5	6	7	8	10	11
8·4	2·1282	1294	1306	1318	1330	1342	1353	1365	1377	1389	1	2	4	5	6	7	8	9	11
8·5	2·1401	1412	1424	1436	1448	1459	1471	1483	1494	1506	1	2	4	5	6	7	8	9	11
8·6	2·1518	1529	1541	1552	1564	1576	1587	1599	1610	1622	1	2	3	5	6	7	8	9	10
8·7	2·1633	1645	1656	1668	1679	1691	1702	1713	1725	1736	1	2	3	5	6	7	8	9	10
8·8	2·1748	1759	1770	1782	1793	1804	1815	1827	1838	1849	1	2	3	5	6	7	8	9	10
8·9	2·1861	1872	1883	1894	1905	1917	1928	1939	1950	1961	1	2	3	4	6	7	8	9	10
9·0	2·1972	1983	1994	2006	2017	2028	2039	2050	2061	2072	1	2	3	4	6	7	8	9	10
9·1	2·2083	2094	2105	2116	2127	2138	2148	2159	2170	2181	1	2	3	4	5	7	8	9	10
9·2	2·2192	2203	2214	2225	2235	2246	2257	2268	2279	2289	1	2	3	4	5	6	8	9	10
9·3	2·2300	2311	2322	2332	2343	2354	2364	2375	2386	2396	1	2	3	4	5	6	7	9	10
9·4	2·2407	2418	2428	2439	2450	2460	2471	2481	2492	2502	1	2	3	4	5	6	7	8	10
9·5	2·2513	2523	2534	2544	2555	2565	2576	2586	2597	2607	1	2	3	4	5	6	7	8	9
9·6	2·2618	2628	2638	2649	2659	2670	2680	2690	2701	2711	1	2	3	4	5	6	7	8	9
9·7	2·2721	2732	2742	2752	2762	2773	2783	2793	2803	2814	1	2	3	4	5	6	7	8	9
9·8	2·2824	2834	2844	2854	2865	2875	2885	2899	2905	2915	1	2	3	4	5	6	7	8	9
9·9	2·2925	2935	2946	2956	2966	2976	2986	2996	3006	3016	1	2	3	4	5	6	7	8	9
10·0	2·3026																		

Hyperbolic or Naperian Logarithms of 10^{-n}.

n	1	2	3	4	5	6	7	8	9
$\log_e 10^{-n}$	$\bar{3}\cdot6974$	$\bar{5}\cdot3948$	$\bar{7}\cdot0922$	$\overline{10}\cdot7897$	$\overline{12}\cdot4871$	$\overline{14}\cdot1845$	$\overline{17}\cdot8819$	$\overline{19}\cdot5793$	$\overline{21}\cdot2767$

EXPONENTI

x	e^x	e^{-x}	sinh x
·02	1·0202	·9802	·0200
·04	1·0408	·9608	·0400
·06	1·0618	·9418	·0600
·08	1·0833	·9231	·0801
·10	1·1052	·9048	·1002
·11	1·1163	·8958	·1102
·12	1·1275	·8869	·1203
·13	1·1388	·8781	·1304
·14	1·1503	·8694	·1405
·15	1·1618	·8607	·1506
·16	1·1735	·8521	·1607
·17	1·1853	·8437	·1708
·18	1·1972	·8353	·1810
·19	1·2092	·8270	·1911
·20	1·2214	·8187	·2013
·21	1·2337	·8106	·2115
·22	1·2461	·8025	·2218
·23	1·2586	·7945	·2320
·24	1·2712	·7866	·2423
·25	1·2840	·7788	·2526
·26	1·2969	·7711	·2629
·27	1·3100	·7634	·2733
·28	1·3231	·7558	·2837
·29	1·3364	·7483	·2941
·30	1·3499	·7408	·3045
·31	1·3634	·7335	·3150
·32	1·3771	·7261	·3255
·33	1·3910	·7189	·3360
·34	1·4050	·7118	·3466
·35	1·4191	·7047	·3572
·36	1·4333	·6977	·3678
·37	1·4477	·6907	·3785
·38	1·4623	·6839	·3892
·39	1·4770	·6771	·4000
·40	1·4917	·6703	·4107
·41	1·5068	·6636	·4216
·42	1·5220	·6570	·4325
·43	1·5373	·6505	·4434
·44	1·5527	·6440	·4543
·45	1·5683	·6376	·4653
·46	1·5841	·6313	·4764
·47	1·6000	·6250	·4875
·48	1·6161	·6188	·4986
·49	1·6323	·6126	·5098
·50	1·6487	·6065	·5211
·6	1·8221	·5488	·6367
·7	2·0138	·4966	·7586
·8	2·2255	·4493	·8881
·9	2·4596	·4066	1·0264

cosh x =

GARITHMS

	Mean Differences								
	1	2	3	4	5	6	7	8	9
	2	4	5	7	9	11	13	14	16
	2	4	5	7	9	11	12	14	16
	2	3	5	7	9	10	12	14	16
	2	3	5	7	9	10	12	14	15
	2	3	5	7	8	10	12	13	15
5	2	3	5	7	8	10	12	13	15
9	2	3	5	6	8	10	11	13	15
0	2	3	5	6	8	10	11	13	14
7	2	3	5	6	8	9	11	13	14
3	2	3	5	6	8	9	11	12	14
5	2	3	5	6	8	9	11	12	14
5	2	3	5	6	8	9	11	12	14
5	1	3	4	6	7	9	10	12	13
1	1	3	4	6	7	9	10	12	13
5	1	3	4	6	7	9	10	12	13
7	1	3	4	6	7	9	10	11	13
7	1	3	4	6	7	8	10	11	13
5	1	3	4	6	7	8	10	11	12
1	1	3	4	5	7	8	10	11	12
6	1	3	4	5	7	8	9	11	12
8	1	3	4	5	7	8	9	11	12
9	1	3	4	5	7	8	9	10	12
8	1	3	4	5	6	8	9	10	12
6	1	3	4	5	6	8	9	10	11
2	1	3	4	5	6	8	9	10	11
6	1	3	4	5	6	7	9	10	11
9	1	2	4	5	6	7	9	10	11
0	1	2	4	5	6	7	9	10	11
0	1	2	4	5	6	7	8	10	11
9	1	2	4	5	6	7	8	9	11
6	1	2	4	5	6	7	8	9	11
2	1	2	3	5	6	7	8	9	10
6	1	2	3	5	6	7	8	9	10
9	1	2	3	5	6	7	8	9	10
1	1	2	3	4	6	7	8	9	10
2	1	2	3	4	6	7	8	9	10
1	1	2	3	4	5	7	8	9	10
9	1	2	3	4	5	6	8	9	10
6	1	2	3	4	5	6	7	9	10
2	1	2	3	4	5	6	7	8	10
7	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9
4	1	2	3	4	5	6	7	8	9
5	1	2	3	4	5	6	7	8	9
6	1	2	3	4	5	6	7	8	9

of 10^{-n}.

	7	8	9
45	$\overline{17}\cdot 8819$	$\overline{19}\cdot 5793$	$\overline{21}\cdot 2767$

Table VII

EXPONENTIAL AND HYPERBOLIC FUNCTIONS

x	e^x	e^{-x}	sinh x	cosh x	x	e^x	e^{-x}	sinh x	cosh x
·02	1·0202	·9802	·0200	1·0002	1·0	2·7183	·3679	1·1752	1·5431
·04	1·0408	·9608	·0400	1·0008	1·1	3·0042	·3329	1·3356	1·6685
·06	1·0618	·9418	·0600	1·0018	1·2	3·3201	·3012	1·5095	1·8107
·08	1·0833	·9231	·0801	1·0032	1·3	3·6693	·2725	1·6984	1·9709
·10	1·1052	·9048	·1002	1·0050	1.4	4.0552	·2466	1.9043	2.1509
·11	1·1163	·8958	·1102	1·0061	1·5	4·4817	·2231	2·1293	2·3524
·12	1·1275	·8869	·1203	1·0072	1·6	4·9530	·2019	2·3756	2·5775
·13	1·1388	·8781	·1304	1·0085	1·7	5·4739	·1827	2·6456	2·8283
·14	1·1503	·8694	·1405	1·0098	1·8	6·0497	·1653	2·9422	3·1075
·15	1·1618	·8607	·1506	1·0113	1.9	6.6859	.1496	3.2682	3.4177
·16	1·1735	·8521	·1607	1·0128	2·0	7·3891	·1353	3·6269	3·7622
·17	1·1853	·8437	·1708	1·0145	2·1	8·1662	·1225	4·0219	4·1443
·18	1·1972	·8353	·1810	1·0162	2·2	9·0250	·1108	4·4571	4·5679
·19	1·2092	·8270	·1911	1·0181	2·3	9·9742	·1003	4·9370	5·0372
·20	1·2214	·8187	·2013	1·0201	2.4	11.023	.0907	5.4662	5.5569
·21	1·2337	·8106	·2115	1·0221	2·5	12·182	·0821	6·0502	6·1323
·22	1·2461	·8025	·2218	1·0243	2·6	13·464	·0743	6·6947	6·7690
·23	1·2586	·7945	·2320	1·0266	2·7	14·880	·0672	7·4063	7·4735
·24	1·2712	·7866	·2423	1·0289	2·8	16·445	·0608	8·1919	8·2527
·25	1·2840	·7788	·2526	1·0314	2.9	18.174	.0550	9.0596	9.1146
·26	1·2969	·7711	·2629	1·0340	3·0	20·085	·0498	10·018	10·068
·27	1·3100	·7634	·2733	1·0367	3·1	22·198	·0450	11·076	11·121
·28	1·3231	·7558	·2837	1·0395	3·2	24·532	·0408	12·246	12·287
·29	1·3364	·7483	·2941	1·0423	3·3	27·113	·0369	13·538	13·575
·30	1·3499	·7408	·3045	1·0453	3.4	29.964	.0334	14.965	14.999
·31	1·3634	·7335	·3150	1·0484	3·5	33·115	·0302	16·543	16·573
·32	1·3771	·7261	·3255	1·0516	3·6	36·598	·0273	18·285	18·313
·33	1·3910	·7189	·3360	1·0550	3·7	40·447	·0247	20·211	20·236
·34	1·4050	·7118	·3466	1·0584	3·8	44·701	·0224	22·339	22·362
·35	1·4191	·7047	·3572	1·0619	3.9	49.402	.0202	24.691	24.711
·36	1·4333	·6977	·3678	1·0655	4·0	54·598	·0183	27·290	27·308
·37	1·4477	·6907	·3785	1·0692	4·1	60·340	·0166	30·162	30·178
·38	1·4623	·6839	·3892	1·0731	4·2	66·686	·0150	33·336	33·351
·39	1·4770	·6771	·4000	1·0770	4·3	73·700	·0136	36·843	36·857
·40	1·4917	·6703	·4107	1·0811	4.4	81.451	.0123	40.719	40.732
·41	1·5068	·6636	·4216	1·0852	4·5	90·017	·0111	45·003	45·014
·42	1·5220	·6570	·4325	1·0895	4·6	99·484	·0100	49·737	49·747
·43	1·5373	·6505	·4434	1·0939	4·7	109·95	·00910	54·969	54·978
·44	1·5527	·6440	·4543	1·0984	4·8	121·51	·00823	60·751	60·759
·45	1·5683	·6376	·4653	1·1030	4.9	134.29	.00745	67.141	67.149
·46	1·5841	·6313	·4764	1·1077	5·0	148·41	·00674	74·203	74·210
·47	1·6000	·6250	·4875	1·1125	5·1	164·02	·00610	82·008	82·014
·48	1·6161	·6188	·4986	1·1174	5·2	181·27	·00552	90·633	90·639
·49	1·6323	·6126	·5098	1·1225	5·3	200·34	·00499	100·17	100·17
·50	1·6487	·6065	·5211	1·1276	5.4	221.41	.00452	110.70	110·71
·6	1·8221	·5488	·6367	1·1855	5·5	244·69	·00409	122·34	122·35
·7	2·0138	·4966	·7586	1·2552	5·6	270·43	·00370	135·21	135·21
·8	2·2255	·4493	·8881	1·3374	5·7	298·87	·00335	149·43	149·43
·9	2·4596	·4066	1·0264	1·4331	5·8	330·30	·00303	165·15	165·15
					5·9	365·04	·00274	182·52	182·52
					6·0	403·43	·00248	201·71	201·72

$$\cosh x = \frac{1}{2}(e^x + e^{-x}),\ \sinh x = \frac{1}{2}(e^x - e^{-x})$$

Table VIII

POWERS, ROOTS AND RECIPROCALS

n	n^2	n^3	$\sqrt{n}$	$\sqrt[3]{n}$	$\sqrt{10n}$	$\sqrt[3]{10n}$	$\sqrt[3]{100n}$	$\frac{1}{n}$
1	1	1	1	1	3·162	2·154	4·642	1
2	4	8	1·414	1·260	4·472	2·714	5·848	·5000
3	9	27	1·732	1·442	5·477	3·107	6·694	·3333
4	16	64	2	1·587	6·325	3·420	7·368	·2500
5	25	125	2·236	1·710	7·071	3·684	7·937	·2000
6	36	216	2·449	1·817	7·746	3·915	8·434	·1667
7	49	343	2·646	1·913	8·367	4·121	8·879	·1429
8	64	512	2·828	2·000	8·944	4·309	9·283	·1250
9	81	729	3·000	2·080	9·487	4·481	9·655	·1111
10	100	1000	3·162	2·154	10·0	4·642	10·000	·1000
11	121	1331	3·317	2·224	10·488	4·791	10·323	·09091
12	144	1728	3·464	2·289	10·994	4·932	10·627	·08333
13	169	2197	3·606	2·351	11·402	5·066	10·914	·07692
14	196	2744	3·742	2·410	11·852	5·192	11·187	·07143
15	225	3375	3·873	2·466	12·247	5·313	11·447	·06667
16	256	4096	4·000	2·520	12·649	5·429	11·696	·06250
17	289	4913	4·123	2·571	13·038	5·540	11·935	·05882
18	324	5832	4·243	2·621	13·416	5·646	12·164	·05556
19	361	6859	4·359	2·668	13·784	5·749	12·386	·05263
20	400	8000	4·472	2·714	14·142	5·848	12·599	·0500
21	441	9261	4·583	2·759	14·491	5·944	12·806	·04762
22	484	10648	4·690	2·802	14·832	6·037	13·006	·04545
23	529	12167	4·796	2·844	15·166	6·127	13·200	·04348
24	576	13824	4·899	2·884	15·492	6·214	13·389	·04167
25	625	15625	5·000	2·924	15·811	6·300	13·572	·0400
26	676	17576	5·099	2·952	16·125	6·383	13·751	·03846
27	729	19683	5·196	3·000	16·432	6·463	13·925	·03704
28	784	21952	5·292	3·037	16·733	6·542	14·095	·03571
29	841	24389	5·385	3·072	17·029	6·619	14·260	·03448
30	900	27000	5·477	3·107	17·321	6·694	14·422	·03333
31	961	29791	5·568	3·141	17·607	6·758	16·581	·03226
32	1024	32768	5·657	3·175	17·889	6·840	14·736	·03125
33	1089	35937	5·745	3·208	18·166	6·910	14·888	·03030
34	1156	39304	5·831	3·240	18·439	6·980	15·037	·02941
35	1225	42875	5·916	3·271	18·708	7·047	15·183	·02857
36	1296	46656	6·000	3·302	18·974	7·114	15·326	·02778
37	1369	50653	6·083	3·332	19·235	7·179	15·467	·02703
38	1444	54872	6·164	3·362	19·494	7·243	15·605	·02632
39	1521	59319	6·245	3·391	19·748	7·306	15·741	·02564
40	1600	64000	6·325	3·420	20·00	7·368	15·874	·0250
41	1681	68921	6·403	3·448	20·248	7·429	16·005	·02439
42	1764	74088	6·481	3·476	20·494	7·489	16·134	·02381
43	1849	79507	6·577	3·503	20·736	7·548	16·261	·02326
44	1936	85184	6·633	3·530	20·976	7·606	16·386	·02273
45	2025	91125	6·708	3·557	21·213	7·663	16·510	·02222
46	2116	97336	6·782	3·583	21·448	7·719	16·631	·02174
47	2209	103823	6·856	3·609	21·679	7·775	16·751	·02128
48	2304	110592	6·928	3·634	21·909	7·830	16·869	·02083
49	2401	117649	7·000	3·659	22·136	7·884	16·983	·02041
50	2500	125000	7·071	3·684	23·361	7·937	17·100	·020

POWE

n	n^2	n^3	
51	2601	132651	
52	2704	140608	
53	2809	148877	
54	2916	157464	
55	3025	166375	
56	3136	175616	
57	3249	185193	
58	3364	195112	
59	3481	205379	
60	3000	216000	
61	3721	226981	
62	3844	238328	
63	3969	250047	
64	4096	262144	
65	4225	274625	
66	4356	287496	
67	4489	300763	
68	4624	314432	
69	4761	328509	
70	4900	343000	
71	5041	357911	
72	5184	373248	
73	5329	389017	
74	5476	405224	
75	5625	421825	
76	5776	438976	
77	5929	456533	
78	6084	474552	
79	6241	493039	
80	6400	512000	
81	6561	531441	
82	6724	551368	
83	6889	571787	
84	7056	592704	
85	7225	614125	
86	7396	636036	
87	7569	658503	
88	7744	681472	
89	7921	704969	
90	8100	729000	
91	8281	753571	
92	8464	778688	
93	8649	804357	
94	8836	830584	
95	9025	857375	
96	9216	884736	
97	9409	912673	
98	9604	941192	
99	9801	970299	
100	10000	1000000	1·

OCALS

$\sqrt[3]{10n}$	$\sqrt[3]{100n}$	$\frac{1}{n}$
2·154	4·642	1
2·714	5·848	·5000
3·107	6·694	·3333
3·420	7·368	·2500
3·684	7·937	·2000
3·915	8·434	·1667
4·121	8·879	·1429
4·309	9·283	·1250
4·481	9·655	·1111
4·642	10·000	·1000
4·791	10·323	·09091
4·932	10·627	·08333
5·066	10·914	·07692
5·192	11·187	·07143
5·313	11·447	·06667
5·429	11·696	·06250
5·540	11·935	·05882
5·646	12·164	·05556
5·749	12·386	·05263
5·848	12·599	·0500
5·944	12·806	·04762
6·037	13·006	·04545
6·127	13·200	·04348
6·214	13·389	·04167
6·300	13·572	·0400
6·383	13·751	·03846
6·463	13·925	·03704
6·542	14·095	·03571
6·619	14·260	·03448
6·694	14·422	·03333
6·758	16·581	·03226
6·840	14·736	·03125
6·910	14·888	·03030
6·980	15·037	·02941
7·047	15·183	·02857
7·114	15·326	·02778
7·179	15·467	·02703
7·243	15·605	·02632
7·306	15·741	·02564
7·368	15·874	·0250
7·429	16·005	·02439
7·489	16·134	·02381
7·548	16·261	·02326
7·606	16·386	·02273
7·663	16·510	·02222
7·719	16·631	·02174
7·775	16·751	·02128
7·830	16·869	·02083
7·884	16·983	·02041
7·937	17·100	·020

Table IX

POWERS, ROOTS AND RECIPROCALS

n	n^2	n^3	$\sqrt{n}$	$\sqrt[3]{n}$	$\sqrt{10n}$	$\sqrt[3]{10n}$	$\sqrt[3]{100n}$	$\frac{1}{n}$
51	2601	132651	7·141	3·708	22·583	7·900	17·213	·01961
52	2704	140608	7·211	3·733	22·804	8·041	17·325	·01923
53	2809	148877	7·280	3·756	23·022	8·093	17·435	·01887
54	2916	157464	7·348	3·780	23·238	8·143	17·544	·01852
55	3025	166375	7·416	3·803	23·452	8·193	17·652	·01818
56	3136	175616	7·483	3·826	23·664	8·243	17·758	·01786
57	3249	185193	7·550	3·849	23·875	8·291	17·863	·01754
58	3364	195112	7·616	3·871	24·083	8·340	17·967	·01724
59	3481	205379	7·681	3·893	24·290	8·387	18·070	·01695
60	3000	216000	7·746	3·915	24·495	8·434	18·171	·01667
61	3721	226981	7·810	3·936	24·698	8·481	18·272	·01639
62	3844	238328	7·874	3·958	24·900	8·527	18·371	·01613
63	3969	250047	7·937	3·979	25·100	8·573	18·469	·01587
64	4096	262144	8·000	4·000	25·298	8·618	18·566	·01562
65	4225	274625	8·062	4·021	25·495	8·662	18·663	·01538
66	4356	287496	8·124	4·041	25·690	8·707	18·758	·01515
67	4489	300763	8·185	4·062	25·884	8·750	18·852	·01493
68	4624	314432	8·246	4·082	26·077	8·794	18·945	·01471
69	4761	328509	8·307	4·102	26·268	8·837	19·038	·01449
70	4900	343000	8·367	4·121	26·458	8·879	19·129	·01429
71	5041	357911	8·426	4·141	26·646	8·921	19·220	·01408
72	5184	373248	8·485	4·160	26·833	8·963	19·310	·01389
73	5329	389017	8·544	4·179	27·019	9·004	19·399	·01370
74	5476	405224	8·602	4·198	27·203	9·045	19·487	·01351
75	5625	421825	8·660	4·217	27·386	9·086	19·574	·01333
76	5776	438976	8·718	4·236	27·568	9·126	19·661	·01316
77	5929	456533	8·775	4·254	27·749	9·166	19·747	·01299
78	6084	474552	8·832	4·273	27·928	9·205	19·832	·01282
79	6241	493039	8·888	4·291	28·107	9·244	19·916	·01266
80	6400	512000	8·944	4·309	28·284	9·283	20·000	·01250
81	6561	531441	9·000	4·327	28·460	9·322	20·083	·01235
82	6724	551368	9·055	4·344	28·646	9·360	20·165	·01220
83	6889	571787	9·110	4·362	28·810	9·398	20·247	·01205
84	7056	592704	9·165	4·380	28·983	9·435	20·328	·01190
85	7225	614125	9·220	4·397	29·155	9·473	20·408	·01176
86	7396	636036	9·274	4·414	29·326	9·510	20·488	·01163
87	7569	658503	9·327	4·431	29·496	9·546	20·567	·01149
88	7744	681472	9·381	4·448	29·665	9·583	20·646	·01136
89	7921	704969	9·434	4·465	29·833	9·619	20·724	·01124
90	8100	729000	9·487	4·481	30·000	9·655	20·801	·01111
91	8281	753571	9·539	4·498	30·166	9·691	20·878	·01099
92	8464	778688	9·592	4·514	30·332	9·726	20·954	·01087
93	8649	804357	9·644	4·531	30·496	9·761	21·029	·01075
94	8836	830584	9·695	4·547	30·659	9·796	21·105	·01064
95	9025	857375	9·747	4·563	30·822	9·830	21·179	·01053
96	9216	884736	9·798	4·579	30·984	9·865	21·253	·01042
97	9409	912673	9·849	4·595	31·145	9·899	21·327	·01031
98	9604	941192	9·899	4·610	31·305	9·933	21·400	·01020
99	9801	970299	9·950	4·626	31·464	9·967	21·472	·01010
100	10000	1000000	10·000	4·642	31·623	10·000	21·544	·0100

Table X

u	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
0·0	·0000	·0040	·0080	·0120	·01595	·0199	·0239	·0279	·0319	·0359
0·1	·0398	·0438	·0478	·0517	·0557	·0596	·0636	0675	·0714	·07535
0·2	·0793	·0832	·0871	·0910	·0948	·0987	·1026	·1064	·1103	·1141
0·3	·1179	·1217	·1255	·1293	·1331	·1368	·1406	·1443	·1480	·1517
0·4	·1554	·1591	·1628	·1664	·1700	·1736	·1772	·1808	·1844	·1879
0·5	·1915	·1950	·1985	·2019	·2054	·2088	·2123	·2157	·2190	·2224
0·6	·22575	·2291	·2324	·2357	·2389	·2422	·2454	·2486	·2518	·2549
0·7	·2580	·26115	·2642	·2673	·2704	·2734	·2764	·2794	·2823	·2852
0·8	·2881	·2910	·2939	·2967	·29955	·3023	·3051	·3078	·3106	·3133
0·9	·3159	·3186	·3212	·3238	·3264	·3289	·3315	·3340	·3365	·3389
1·0	·3413	·3438	·3461	·3485	·3508	·3531	·3554	·3577	·3599	·3621
1·1	·3643	·3665	·3686	·3708	·3729	·3749	·3770	·3790	·3810	·3830
1·2	·3849	·3869	·3888	·3907	·3925	·3944	·3962	·3980	·3997	·4015
1·3	·4032	·4049	·4066	·4082	·4099	·4115	·4131	·4147	·4162	·4177
1·4	·4192	·4207	·4222	·4236	·4251	·4265	·4279	·4292	·4306	·4319
1·5	·4332	·4345	·4357	·4370	·4382	·4394	·4406	·4418	·4429	·4441
1·6	·4452	·4463	·4474	·44845	·4495	·4505	·4515	·4525	·4535	·4545
1·7	·4554	·4564	·4573	·4582	·4591	·4599	·4608	·4616	·4625	·4633
1·8	·4641	·4649	·4656	·4664	·4671	·4678	·4686	·4693	·4699	·4706
1·9	·4713	·4719	·4726	·4732	·4738	·4744	·4750	·4756	·4761	·4767
2·0	·4772	·4778	·4783	·4788	·4793	·4798	·4803	·4808	·4812	·4817
2·1	·4821	·4826	·4830	·4834	·4838	·4842	·4846	·4850	·4854	·4857
2·2	·4861	·48645	·4868	·4871	·4875	·4878	·4881	·4884	·4887	·4890
2·3	·4893	·4896	·4898	·4901	·4904	·4906	·4909	·4911	·4913	·4916
2·4	·4918	·4920	·4922	·4925	·4927	·4929	·4931	·4932	·4934	·4936
2·5	·4938	·4940	·4941	·4943	·4945	·4946	·4948	·4949	·4951	·4952
2·6	·4953	·4955	·4956	·4957	·4959	·4960	·4961	·4962	·4963	·4964
2·7	·4965	·4966	·4967	·4968	·4969	·4970	·4971	·4972	·4973	·4974
2·8	·4974	·4975	·4976	·4977	·4977	·4978	·4979	·49795	·4980	·4981
2·9	·4981	·4982	·49825	·4983	·4984	·4984	·4985	·4985	·4986	·4986
3·0	·49865	·4987	·4987	·4988	·4988	·4989	·4989	·4989	·4990	·4990
3·1	·4990	·4991	·4991	·4991	·4992	·4992	·4992	·4992	·4993	·4993
3·2	·4993	·4993	·4994	·4994	·4994	·4994	·4994	·4995	·4995	·4995
3·3	·4995	·4995	·4995	·4996	·4996	·4996	·4996	·4996	·4996	·4997
3·4	·4997	·4997	·4997	·4997	·4997	·4997	·4997	·4997	·49975	·4998
3·5	·4998	·4998	·4998	·4998	·4998	·4998	·4998	·4998	·4998	·4998
3·6	·4998	·4998	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999
3·7	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·4999
3·8	·4999	·4999	·4999	·4999	·4999	·4999	·4999	·49995	·49995	·49995
3·9	·49995	·49995	·5000	·5000	·5000	·5000	·5000	·5000	·5000	·5000

x	0·00	0·01	0·02
0·0	·3989	·3989	·3989
0·1	·3970	·3965	·3961
0·2	·3910	·3902	·3894
0·3	·3814	·3802	·3790
0·4	·3683	·3668	·3653
0·5	·3521	·3503	·3485
0·6	·3332	·3312	·3292
0·7	·3123	·3101	·3079
0·8	·2897	·2874	·2850
0·9	·2661	·2637	·2613
1·0	·2420	·2396	·2371
1·1	·2179	·2155	·2131
1·2	·1942	·1919	·1895
1·3	·1714	·1691	·1669
1·4	·1497	·1476	·1456
1·5	·1295	·1276	·1257
1·6	·1109	·1092	·1074
1·7	·0940	·0925	·0909
1·8	·0790	·0775	·0761
1·9	·0656	·0644	·0632
2·0	·0540	·0529	·0519
2·1	·0440	·0431	·0422
2·2	·0355	·0347	·0339
2·3	·0283	·0277	·0270
2·4	·0224	·0219	·0213
2·5	·0175	·0171	·0167
2·6	·0136	·0132	·0129
2·7	·0104	·0101	·0099
2·8	·0079	·0077	·0075
2·9	·0060	·0058	·0056
3·0	·0044	·0043	·0042
3·1	·0033	·0032	·0031
3·2	·0024	·0023	·0022
3·3	·0017	·0017	·0016
3·4	·0012	·0012	·0012
3·5	·0009	·0008	·0008
3·6	·0006	·0006	·0006
3·7	·0004	·0004	·0004
3·8	·0003	·0003	·0003
3·9	·0002	·0002	·0002

5	0·07	0·08	0·09
39	·0279	·0319	·0359
36	0675	·0714	·07535
26	·1064	·1103	·1141
06	·1443	·1480	·1517
72	·1808	·1844	·1879
23	·2157	·2190	·2224
54	·2486	·2518	·2549
54	·2794	·2823	·2852
51	·3078	·3106	·3133
5	·3340	·3365	·3389
54	·3577	·3599	·3621
70	·3790	·3810	·3830
52	·3980	·3997	·4015
31	·4147	·4162	·4177
79	·4292	·4306	·4319
06	·4418	·4429	·4441
15	·4525	·4535	·4545
08	·4616	·4625	·4633
36	·4693	·4699	·4706
50	·4756	·4761	·4767
03	·4808	·4812	·4817
46	·4850	·4854	·4857
81	·4884	·4887	·4890
09	·4911	·4913	·4916
31	·4932	·4934	·4936
48	·4949	·4951	·4952
51	·4962	·4963	·4964
71	·4972	·4973	·4974
79	·49795	·4980	·4981
85	·4985	·4986	·4986
89	·4989	·4990	·4990
92	·4992	·4993	·4993
94	·4995	·4995	·4995
96	·4996	·4996	·4997
97	·4997	·49975	·4998
98	·4998	·4998	·4998
99	·4999	·4999	·4999
99	·4999	·4999	·4999
99	·49995	·49995	·49995
00	·5000	·5000	·5000

Table XI

x	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
0·0	·3989	·3989	·3989	·3989	·3986	·3984	·3982	·3980	·3977	·3973
0·1	·3970	·3965	·3961	·3956	·3951	·3945	·3939	·3932	·3925	·3918
0·2	·3910	·3902	·3894	·3885	·3876	·3867	·3857	·3847	·3836	·3825
0·3	·3814	·3802	·3790	·3778	·3765	·3752	·3739	·3725	·3712	·3697
0·4	·3683	·3668	·3653	·3637	·3621	·3605	·3589	·3572	·3555	·3538
0·5	·3521	·3503	·3485	·3467	·3448	·3429	·3410	·3391	·3372	·3352
0·6	·3332	·3312	·3292	·3271	·3251	·3230	·3209	·3187	·3166	·3144
0·7	·3123	·3101	·3079	·3056	·3034	·3011	·2989	·2966	·2943	·2920
0·8	·2897	·2874	·2850	·2827	·2803	·2780	·2756	·2732	·2709	·2685
0·9	·2661	·2637	·2613	·2589	·2565	·2541	·2516	·2492	·2468	·2444
1·0	·2420	·2396	·2371	·2347	·2323	·2299	·2275	·2251	·2227	·2203
1·1	·2179	·2155	·2131	·2107	·2083	·2059	·2036	·2012	·1989	·1965
1·2	·1942	·1919	·1895	·1872	·1849	·1826	·1804	·1781	·1758	·1736
1·3	·1714	·1691	·1669	·1647	·1626	·1604	·1582	·1561	·1539	·1518
1·4	·1497	·1476	·1456	·1435	·1415	·1394	·1374	·1354	·1334	·1315
1·5	·1295	·1276	·1257	·1238	·1219	·1200	·1182	·1163	·1145	·1127
1·6	·1109	·1092	·1074	·1057	·1040	·1023	·1006	·0989	·0973	·0957
1·7	·0940	·0925	·0909	·0893	·0878	·0863	·0848	·0833	·0818	·0804
1·8	·0790	·0775	·0761	·0748	·0734	·0721	·0707	·0694	·0681	·0669
1·9	·0656	·0644	·0632	·0620	·0608	·0596	·0584	·0573	·0562	·0551
2·0	·0540	·0529	·0519	·0508	·0498	·0488	·0478	·0468	·0459	·0449
2·1	·0440	·0431	·0422	·0413	·0404	·0396	·0387	·0379	·0371	·0363
2·2	·0355	·0347	·0339	·0332	·0325	·0317	·0310	·0303	·0297	·0290
2·3	·0283	·0277	·0270	·0264	·0258	·0252	·0246	·0241	·0235	·0229
2·4	·0224	·0219	·0213	·0208	·0203	·0198	·0194	·0189	·0184	·0180
2·5	·0175	·0171	·0167	·0163	·0158	·0154	·0151	·0147	·0143	·0139
2·6	·0136	·0132	·0129	·0126	·0122	·0119	·0116	·0113	·0110	·0107
2·7	·0104	·0101	·0099	·0096	·0093	·0091	·0088	·0086	·0084	·0081
2·8	·0079	·0077	·0075	·0073	·0071	·0069	·0067	·0065	·0063	·0061
2·9	·0060	·0058	·0056	·0055	·0053	·0051	·0050	·0048	·0047	·0046
3·0	·0044	·0043	·0042	·0040	·0039	·0038	·0037	·0036	·0035	·0034
3·1	·0033	·0032	·0031	·0030	·3329	·0028	·0027	·0026	·0025	·0025
3·2	·0024	·0023	·0022	·0022	·0021	·0020	·0020	·0019	·0018	·0018
3·3	·0017	·0017	·0016	·0016	·0015	·0015	·0014	·0014	·0013	·0013
3·4	·0012	·0012	·0012	·0011	·0011	·0010	·0010	·0010	·0009	·0009
3·5	·0009	·0008	·0008	·0008	·0008	·0007	·0007	·0007	·0007	·0006
3·6	·0006	·0006	·0006	·0005	·0005	·0005	·0005	·0005	·0005	·0004
3·7	·0004	·0004	·0004	·0004	·0004	·0004	·0003	·0003	·0003	·0003
3·8	·0003	·0003	·0003	·0003	·0003	·0002	·0002	·0002	·0002	·0002
3·9	·0002	·0002	·0002	·0002	·0002	·0002	·0002	·0002	·0001	·0001

Table XII

POBABILITY

r	·99	·95	·90	·70	·50	·30	·10	·05	·01
1	·08157	·00393	·0158	·143	·455	1·074	2·706	3·841	6·635
2	·0201	·103	·211	·713	1·386	2·408	4·605	5·991	9·210
3	·115	·352	·584	1·424	2·366	3·665	6·251	7·815	11·345
4	·297	·711	1·064	2·195	3·357	4·878	7·779	9·488	13·277
5	·554	1·145	1·610	3·000	4·351	6·064	9·236	11·070	15·086
6	·872	1·635	2·204	3·828	5·348	7·231	10·645	12·592	16·812
7	1·239	2·167	2·833	4·671	6·346	8·383	12·017	14·067	18·475
8	1·646	2·733	3·490	5·527	7·344	9·524	13·362	15·507	20·090
9	2·088	3·325	4·168	6·393	8·343	10·656	14·684	16·919	21·666
10	2·558	3·940	4·865	7·267	9·342	11·781	15·987	18·307	23·209
11	3·053	4·575	5·578	8·148	10·341	12·899	17·275	19·675	24·725
12	3·571	5·226	6·304	9·034	11·340	14·011	18·549	21·026	26·217
13	4·107	5·892	7·042	9·926	12·340	15·119	19·812	22·362	27·698
14	4·660	6·571	7·790	10·821	13·339	16·222	21·064	23·685	29·141
15	5·229	7·261	8·547	11·721	14·339	17·322	22·307	24·996	30·578
16	5·812	7·962	9·312	12·624	15·338	18·418	23·542	26·296	32·000
17	6·408	8·672	10·085	13·531	16·338	19·511	24·769	27·587	33·409
18	7·015	9·390	10·865	14·440	17·338	20·601	25·989	28·869	34·805
19	7·633	10·117	11·651	15·352	18·338	21·689	27·204	30·144	36·191
20	8·260	10·851	12·443	16·266	19·337	22·775	28·412	31·410	37·566
21	8·897	11·591	13·240	17·182	20·337	23·853	29·615	32·671	38·932
22	9·542	12·338	14·041	18·101	21·337	24·939	30·813	33·924	40·289
23	10·196	13·091	14·848	19·021	22·337	26·018	32·007	35·172	41·638
24	10·856	13·848	15·659	19·943	23·337	27·096	33·196	36·415	42·980
25	11·524	14·611	16·473	20·867	24·337	28·172	34·382	37·652	44·314
26	12·198	15·379	17·292	21·792	25·336	29·246	35·563	38·885	45·642
27	12·879	16·151	18·114	22·719	26·336	30·319	36·741	40·113	46·963
28	13·565	16·928	18·939	23·647	27·336	31·391	37·916	41·337	48·278
29	14·256	17·708	19·768	24·577	28·336	32·461	39·087	42·557	49·588
30	14·953	18·493	20·599	25·508	29·336	33·530	40·256	43·773	50·892

n	·9	·5
1	·158	1·000
2	·142	·816
3	·137	·765
4	·134	·741
5	·132	·727
6	·131	·718
7	·130	·711
8	·130	·706
9	·129	·703
10	·129	·700
11	·129	·697
12	·128	·695
13	·128	·694
14	·128	·692
15	·128	·691
16	·128	·690
17	·128	·689
18	·127	·688
19	·127	·688
20	·127	·687
21	·127	·686
22	·127	·686
23	·127	·685
24	·127	·685
25	·127	·684
26	·127	·684
27	·127	·684
28	·127	·683
29	·127	·683
30	·127	·683
40	·126	·681
60	·126	·679
120	·126	·677
∞	·126	·674

·10	·05	·01
2·706	3·841	6·635
4·605	5·991	9·210
6·251	7·815	11·345
7·779	9·488	13·277
9·236	11·070	15·086
10·645	12·592	16·812
12·017	14·067	18·475
13·362	15·507	20·090
14·684	16·919	21·666
15·987	18·307	23·209
17·275	19·675	24·725
18·549	21·026	26·217
19·812	22·362	27·698
21·064	23·685	29·141
22·307	24·996	30·578
23·542	26·296	32·000
24·769	27·587	33·409
25·989	28·869	34·805
27·204	30·144	36·191
28·412	31·410	37·566
29·615	32·671	38·932
30·813	33·924	40·289
32·007	35·172	41·638
33·196	36·415	42·980
34·382	37·652	44·314
35·563	38·885	45·642
36·741	40·113	46·963
37·916	41·337	48·278
39·087	42·557	49·588
40·256	43·773	50·892

Table XIII

POBABILITY

n	·9	·5	·4	·1	·05	·01
1	·158	1·000	1·376	6·314	12·706	63·657
2	·142	·816	1·061	2·920	4·303	9·925
3	·137	·765	·978	2·353	3·182	5·841
4	·134	·741	·941	2·132	2·776	4·604
5	·132	·727	·920	2·015	2·571	4·032
6	·131	·718	·906	1·943	2·447	3·707
7	·130	·711	·896	1·895	2·365	3·499
8	·130	·706	·889	1·860	2·306	3·355
9	·129	·703	·883	1·833	2·262	3·250
10	·129	·700	·879	1·812	2·228	3·169
11	·129	·697	·876	1·796	2·201	3·106
12	·128	·695	·873	1·782	2·179	3·055
13	·128	·694	·870	1·771	2·160	3·012
14	·128	·692	·868	1·761	2·145	2·977
15	·128	·691	·866	1·753	2·131	2·947
16	·128	·690	·865	1·746	2·120	2·921
17	·128	·689	·863	1·740	2·110	2·898
18	·127	·688	·862	1·734	2·101	2·878
19	·127	·688	·861	1·729	2·093	2·861
20	·127	·687	·860	1·745	2·086	2·845
21	·127	·686	·859	1·721	2·080	2·831
22	·127	·686	·858	1·717	2·074	2·819
23	·127	·685	·858	1·714	2·069	2·807
24	·127	·685	·857	1·711	2·064	2·797
25	·127	·684	·856	1·708	2·060	2·787
26	·127	·684	·856	1·706	2·056	2·779
27	·127	·684	·855	1·703	2·052	2·771
28	·127	·683	·855	1·701	2·048	2·763
29	·127	·683	·854	1·699	2·045	2·756
30	·127	·683	·854	1·697	2·042	2·750
40	·126	·681	·851	1·684	2·021	2·704
60	·126	·679	·848	1·671	2·000	2·660
120	·126	·677	·845	1·658	1·980	2·617
∞	·126	·674	·842	1·645	1·960	2·576

Table XIV (5% points of F)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	∞
1	161·45	199·50	215·71	224·58	230·16	233·99	236·77	238·88	240·54	241·88	243·91	245·95	248·01	249·05	254·32
2	18·51	19·00	19·16	19·25	19·30	19·33	19·35	19·37	19·38	19·40	19·41	19·43	19·45	19·45	19·50
3	10·13	9·55	9·28	9·12	9·01	8·94	8·89	8·85	8·81	8·79	8·74	8·70	8·66	8·64	8·53
4	7·71	6·94	6·59	6·39	6·26	6·16	6·09	6·04	6·00	5·96	5·91	5·86	5·80	5·77	5·63
5	6·61	5·79	5·41	5·19	5·05	4·95	4·88	4·82	4·77	4·74	4·68	4·62	4·56	4·53	4·36
6	5·99	5·14	4·76	4·53	4·39	4·28	4·21	4·15	4·10	4·06	4·00	3·94	3·87	3·84	3·67
7	5·59	4·74	4·35	4·12	3·97	3·87	3·79	3·73	3·68	3·64	3·57	3·51	3·44	3·41	3·23
8	5·32	4·46	4·07	3·84	3·69	3·58	3·50	3·44	3·39	3·35	3·28	3·22	3·15	3·12	2·93
9	5·12	4·26	3·86	3·63	3·48	3·37	3·29	3·23	3·18	3·14	3·07	3·01	2·94	2·90	2·71
10	4·96	4·10	3·71	3·48	3·33	3·22	3·14	3·07	3·02	2·98	2·91	2·84	2·77	2·74	2·54
11	4·84	3·98	3·59	3·36	3·20	3·09	3·01	2·95	2·90	2·85	2·79	2·72	2·65	2·61	2·40
12	4·75	3·89	3·49	3·26	3·11	3·00	2·91	2·85	2·80	2·75	2·69	2·62	2·54	2·51	2·30
13	4·67	3·81	3·41	3·18	3·03	2·92	2·83	2·77	2·71	2·67	2·60	2·53	2·46	2·42	2·21
14	4·60	3·74	3·34	3·11	2·96	2·85	2·76	2·70	2·65	2·60	2·51	2·46	2·39	2·35	2·13
15	4·54	3·68	3·29	3·06	2·90	2·79	2·71	2·64	2·59	2·54	2·48	2·40	2·33	2·29	2·07
16	4·49	3·63	3·24	3·01	2·85	2·74	2·66	2·59	2·54	2·49	2·42	2·35	2·28	2·24	2·01
17	4·45	3·59	3·20	2·96	2·81	2·70	2·61	2·55	2·49	2·45	2·38	2·31	2·23	2·19	1·96
18	4·41	3·55	3·16	2·93	2·77	2·66	2·58	2·51	2·46	2·41	2·34	2·27	2·19	2·15	1·92
19	4·38	3·52	3·13	2·90	2·74	2·63	2·54	2·48	2·42	2·38	2·31	2·23	2·16	2·11	1·88
20	4·35	3·49	3·10	2·87	2·71	2·60	2·51	2·45	2·39	2·35	2·28	2·20	2·12	2·08	1·84
21	4·32	3·47	3·07	2·84	2·68	2·57	2·49	2·42	2·37	2·32	2·25	2·18	2·10	2·05	1·81
22	4·30	3·44	3·05	2·82	2·66	2·55	2·46	2·40	2·34	2·30	2·23	2·15	2·07	2·03	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·44	2·37	2·32	2·27	2·20	2·13	2·05	2·01	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·42	2·36	2·30	2·25	2·18	2·11	2·03	1·98	1·73
25	4·24	3·39	2·99	2·76	2·60	2·49	2·40	2·34	2·28	2·24	2·16	2·09	2·01	1·96	1·71
26	4·23	3·37	2·98	2·74	2·59	2·47	2·39	2·32	2·27	2·22	2·15	2·07	1·99	1·95	1·69
27	4·21	3·35	2·96	2·73	2·57	2·46	2·37	2·31	2·25	2·20	2·13	2·06	1·97	1·93	1·67
28	4·20	3·34	2·95	2·71	2·56	2·45	2·36	2·29	2·24	2·19	2·12	2·04	1·96	1·91	1·65
29	4·18	3·33	2·93	2·70	2·55	2·43	2·35	2·28	2·22	2·18	2·10	2·03	1·94	1·90	1·64
30	4·71	3·32	2·92	2·69	2·53	2·42	2·33	2·27	2·21	2·16	2·09	2·01	1·93	1·89	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·25	2·18	2·12	2·08	2·00	1·92	1·84	1·79	1·51
60	4·00	3·15	2·76	2·53	2·37	2·25	2·17	2·10	2·04	1·99	1·92	1·84	1·75	1·70	1·39
120	3·92	3·07	2·68	2·45	2·29	2·17	2·09	2·02	1·96	1·91	1·83	1·75	1·66	1·61	1·25
∞	3·84	3·00	2·60	2·37	2·21	2·10	2·01	1·94	1·88	1·83	1·75	1·67	1·57	1·52	1·00

Table XV (1% points of F)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	∞
1	4,052·4	4,999·5	5,403·3	5,624·6	5,763·7	5,859·0	5,928·3	5,981·6	6,022·5	6,055·8	6,106·3	6,157·3	6,208·7	6,234·6	6,366·0
2	98·50	99·00	99·17	99·25	99·30	99·33	99·36	99·37	99·39	99·40	99·42	99·43	99·45	99·46	99·50
3	34·12	30·82	29·46	28·71	28·24	27·91	27·67	27·49	27·34	27·23	27·05	26·87	26·69	26·60	26·12
4	21·20	18·00	16·69	15·98	15·52	15·21	14·98	14·80	14·66	14·55	14·37	14·20	14·02	13·93	13·46
5	16·26	13·27	12·06	11·39	10·97	10·67	10·46	10·29	10·16	10·05	9·89	9·72	9·55	9·47	9·02
6	13·74	10·92	9·78	9·15	8·75	8·47	8·26	8·10	7·98	7·87	7·72	7·56	7·40	7·31	6·88
7	12·25	9·55	8·45	7·85	7·46	7·19	6·99	6·84	6·72	6·62	6·47	6·31	6·16	6·07	5·65
8	11·26	8·65	7·59	7·01	6·63	6·37	6·18	6·03	5·91	5·81	5·67	5·52	5·36	5·28	4·86
9	10·56	8·02	6·99	6·42	6·06	5·80	5·61	5·47	5·35	5·26	5·11	4·96	4·81	4·73	4·31
10	10·04	7·56	6·55	5·99	5·64	5·39	5·20	5·06	4·94	4·85	4·71	4·56	4·41	4·33	3·91

22	4·30	3·44	3·05	2·82	2·66	2·55	2·46	2·40	2·34	2·30	2·23	2·15	2·07	2·03	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·44	2·37	2·32	2·27	2·20	2·13	2·05	2·01	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·42	2·36	2·30	2·25	2·18	2·11	2·03	1·98	1·73
25	4·24	3·39	2·99	2·76	2·60	2·49	2·40	2·34	2·28	2·24	2·16	2·09	2·01	1·96	1·71
26	4·23	3·37	2·98	2·74	2·59	2·47	2·39	2·32	2·27	2·22	2·15	2·07	1·99	1·95	1·69
27	4·21	3·35	2·96	2·73	2·57	2·46	2·37	2·31	2·25	2·20	2·13	2·06	1·97	1·93	1·67
28	4·20	3·34	2·95	2·71	2·56	2·45	2·36	2·29	2·24	2·19	2·12	2·04	1·96	1·91	1·65
29	4·18	3·33	2·93	2·70	2·55	2·43	2·35	2·28	2·22	2·18	2·10	2·03	1·94	1·90	1·64
30	4·71	3·32	2·92	2·69	2·53	2·42	2·33	2·27	2·21	2·16	2·09	2·01	1·93	1·89	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·25	2·18	2·12	2·08	2·00	1·92	1·84	1·79	1·51
60	4·00	3·15	2·76	2·53	2·37	2·25	2·17	2·10	2·04	1·99	1·92	1·84	1·75	1·70	1·39
120	3·92	3·07	2·68	2·45	2·29	2·17	2·09	2·02	1·96	1·91	1·83	1·75	1·66	1·61	1·25
∞	3·84	3·00	2·60	2·37	2·21	2·10	2·01	1·94	1·88	1·83	1·75	1·67	1·57	1·52	1·00

Table XV (1% points of F)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	∞
1	4,052·4	4,999·5	5,403·3	5,624·6	5,763·7	5,859·0	5,928·3	5,981·6	6,022·5	6,055·8	6,106·3	6,157·3	6,208·7	6,234·6	6,366·0
2	98·50	99·00	99·17	99·25	99·30	99·33	99·36	99·37	99·39	99·40	99·42	99·43	99·45	99·46	99·50
3	34·12	30·82	29·46	28·71	28·24	27·91	27·67	27·49	27·34	27·23	27·05	26·87	26·69	26·60	26·12
4	21·20	18·00	16·69	15·98	15·52	15·21	14·98	14·80	14·66	14·55	14·37	14·20	14·02	13·93	13·46
5	16·26	13·27	12·06	11·39	10·97	10·67	10·46	10·29	10·16	10·05	9·89	9·72	9·55	9·47	9·02
6	13·74	10·92	9·78	9·15	8·75	8·47	8·26	8·10	7·98	7·87	7·72	7·56	7·40	7·31	6·88
7	12·25	9·55	8·45	7·85	7·46	7·19	6·99	6·84	6·72	6·62	6·47	6·31	6·16	6·07	5·65
8	11·26	8·65	7·59	7·01	6·63	6·37	6·18	6·03	5·91	5·81	5·67	5·52	5·36	5·28	4·86
9	10·56	8·02	6·99	6·42	6·06	5·80	5·61	5·47	5·35	5·26	5·11	4·96	4·81	4·73	4·31
10	10·04	7·56	6·55	5·99	5·64	5·39	5·20	5·06	4·94	4·85	4·71	4·56	4·41	4·33	3·91
11	9·65	7·21	6·22	5·67	5·32	5·07	4·89	4·74	4·63	4·54	4·40	4·25	4·10	4·02	3·60
12	9·33	6·93	5·95	5·41	5·06	4·82	4·64	4·50	4·39	4·30	4·16	4·01	3·86	3·78	3·36
13	9·07	6·70	5·74	5·21	4·86	4·62	4·44	4·30	4·19	4·10	3·96	3·82	3·66	3·59	3·17
14	8·86	6·51	5·56	5·04	4·69	4·46	4·28	4·14	4·03	3·94	3·80	3·66	3·51	3·43	3·00
15	8·68	6·36	5·42	4·89	4·56	4·32	4·14	4·00	3·89	3·80	3·67	3·52	3·37	3·29	2·87
16	8·53	6·23	5·29	4·77	4·44	4·20	4·03	3·89	3·78	3·69	3·55	3·41	3·26	3·18	2·75
17	8·40	6·11	5·18	4·67	4·34	4·10	3·93	3·79	3·68	3·59	3·46	3·31	3·16	3·08	2·65
18	8·29	6·01	5·09	4·58	4·25	4·01	3·84	3·71	3·60	3·51	3·37	3·23	3·08	3·00	2·57
19	8·18	5·93	5·01	4·50	4·17	3·94	3·77	3·63	3·52	3·43	3·30	3·15	3·00	2·92	2·49
20	8·10	5·85	4·94	4·43	4·10	3·87	3·70	3·56	3·46	3·37	3·23	3·09	2·94	2·86	2·42
21	8·0	5·78	4·87	4·37	4·04	3·81	3·64	3·51	3·40	3·31	3·17	3·03	2·88	2·80	2·36
22	7·95	5·72	4·82	4·31	3·99	3·76	3·59	3·45	3·35	3·26	3·12	2·98	2·83	2·75	2·31
23	7·88	5·66	4·76	4·26	3·94	3·71	3·54	3·41	3·30	3·21	3·07	2·93	2·78	2·70	2·26
24	7·82	5·61	4·72	4·22	3·90	3·67	3·50	3·36	3·26	3·17	3·03	2·89	2·74	2·66	2·21
25	7·77	5·57	4·68	4·18	3·86	3·63	3·46	3·32	3·22	3·13	2·99	2·85	2·70	2·62	2·17
26	7·72	5·53	4·64	4·14	3·82	3·59	3·42	3·29	3·18	3·09	2·96	2·81	2·66	2·58	2·13
27	7·68	5·49	4·60	4·11	3·78	3·56	3·39	3·26	3·15	3·06	2·93	2·78	2·63	2·55	2·10
28	7·64	5·45	4·57	4·07	3·75	3·53	3·36	3·23	3·12	3·03	2·90	2·75	2·60	2·52	2·06
29	7·60	5·42	4·54	4·04	3·73	3·50	3·33	3·20	3·09	3·00	2·87	2·73	2·57	2·49	2·03
30	7·56	5·39	4·51	4·02	3·70	3·47	3·30	3·17	3·07	2·98	2·84	2·70	2·55	2·47	2·01
40	7·31	5·18	4·31	3·83	3·51	3·29	3·12	2·99	2·89	2·80	2·66	2·52	2·37	2·29	1·80
60	7·08	4·98	4·13	3·65	3·34	3·12	2·95	2·82	2·72	2·63	2·50	2·35	2·20	2·12	1·60
120	6·85	4·79	3·95	3·48	3·17	2·96	2·79	2·66	2·56	2·47	2·34	2·19	2·03	1·95	1·38
∞	6·63	4·61	3·78	3·32	3·02	2·80	2·64	2·51	2·41	2·32	2·18	2·04	1·88	1·79	1·00

Table XVI

5% POINTS OF Z

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	12	24	∞
1	2·5421	2·6479	2·6870	2·7071	2·7194	2·7276	2·7380	2·7484	2·7588	2·7693
2	1·4592	1·4722	1·4765	1·4787	1·4800	1·4808	1·4819	1·4830	1·4840	1·4851
3	1·1577	1·1284	1·1137	1·1051	1·0994	1·0953	1·0899	1·0842	1·0781	1·0716
4	1·0212	·9690	·9429	·9272	·9168	·9093	·8993	·8885	·8767	·8639
5	·9441	·8777	·8441	·8236	·8097	·7997	·7862	·7714	·7550	·7368
6	·8948	·8188	·7798	·7558	·7394	·7274	·7112	·6931	·6729	·6499
7	·8606	·7777	·7347	·7080	·6896	·6761	·6576	·6369	·6134	·5862
8	·8355	·7475	·7014	·6725	·6525	·6378	·6175	·5945	·5682	·5371
9	·8163	·7242	·6757	·6450	·6238	·6080	·5862	·5613	·5324	·4979
10	·8012	·7058	·6553	·6232	·6009	·5843	·5611	·5346	·5035	·4657
11	·7889	·6909	·6387	·6055	·5822	·5648	·5406	·5126	·4795	·4387
12	·7788	·6786	·6250	·5907	·5666	·5487	·5234	·4941	·4592	·4156
13	·7703	·6682	·6134	·5783	·5535	·5350	·5089	·4785	·4419	·3957
14	·7630	·6594	·6036	·5677	·5423	·5233	·4964	·4649	·4269	·3782
15	·7568	·6518	·5950	·5585	·5326	·5131	·4855	·4532	·4138	·3628
16	·7514	·6451	·5876	·5505	·5241	·5042	·4760	·4428	·4022	·3490
17	·7466	·6393	·5811	·5434	·5166	·4964	·4676	·4337	·3919	·3366
18	·7424	·6341	·5753	·5371	·5099	·4894	·4602	·4255	·3827	·3253
19	·7386	·6295	·5701	·5315	·5040	·4832	·4535	·4182	·3743	·3151
20	·7352	·6254	·5654	·5265	·4986	·4776	·4474	·4116	·3668	·3057
21	·7322	·6216	·5612	·5219	·4938	·4725	·4420	·4055	·3599	·2971
22	·7294	·6182	·5574	·5178	·4894	·4679	·4870	·4001	·3536	·2892
23	·7269	·6151	·5540	·5140	·4854	·4636	·4325	·3950	·3478	·2818
24	·7246	·6123	·5508	·5106	·4817	·4598	·4283	·3904	·3425	·2749
25	·7225	·6097	·5478	·5074	·4783	·4562	·4244	·3862	·3376	·2635
26	·7205	·6073	·5451	·5045	·4752	·4529	·4209	·3823	·3330	·2625
27	·7187	·6051	·5427	·5017	·4723	·4499	·4176	·3786	·3287	·2569
28	·7171	·6030	·5403	·4992	·4696	·4471	·4146	·3752	·3248	·2516
29	·7155	·6011	·5382	·4969	·4671	·4444	·4117	·3720	·3211	·2466
30	·7141	·5994	·5362	·4947	·4648	·4420	·4090	·3691	·3176	·2419
40	·7037	·5866	·5217	·4789	·4479	·4242	·3897	·3475	·2920	·2057
60	·6933	·5738	·5073	·4632	·4311	·4064	·3702	·3255	·2654	·1644
120	·6830	·5611	·4930	·4475	·4143	·3885	·3506	·3032	·2376	·1131
∞	·6729	·5486	·4787	·4319	·3974	·3706	·3309	·2804	·2085	0

$\nu_2 \backslash \nu_1$	1	2	3
1	4·1535	4·2585	4·297
2	2·2950	2·2976	2·298
3	1·7649	1·7140	1·691
4	1·5270	1·4452	1·407
5	1·3973	1·2929	1·244
6	1·3103	1·1955	1·1401
7	1·2526	1·1281	1·0672
8	1·2106	1·0787	1·0135
9	1·1786	1·0411	·9724
10	1·1535	1·0114	·9399
11	1·1333	·9874	·9136
12	1·1166	·9677	·8919
13	1·1027	·9511	·8737
14	1·0909	·9370	·8581
15	1·0807	·9249	·8448
16	1·0719	·9144	·8331
17	1·0641	·9051	·8229
18	1·0572	·8970	·8138
19	1·0511	·8897	·8057
20	1·0457	·8831	·7985
21	1·0408	·8772	·7920
22	1·0363	·8719	·7860
23	1·0322	·8670	·7806
24	1·0285	·8626	·7757
25	1·0251	·8585	·7712
26	1·0220	·8548	·7670
27	1·0191	·8513	·7631
28	1·0164	·8481	·7595
29	1·0139	·8451	·7562
30	1·0116	·8423	·7531
40	·9949	·8223	·7307
60	·9784	·8025	·7086
120	·9622	·7829	·6867
∞	·9462	·7636	·6651

8	12	24	∞
2·7380	2·7484	2·7588	2·7693
1·4819	1·4830	1·4840	1·4851
1·0899	1·0842	1·0781	1·0716
·8993	·8885	·8767	·8639
·7862	·7714	·7550	·7368
·7112	·6931	·6729	·6499
·6576	·6369	·6134	·5862
·6175	·5945	·5682	·5371
·5862	·5613	·5324	·4979
·5611	·5346	·5035	·4657
·5406	·5126	·4795	·4387
·5234	·4941	·4592	·4156
·5089	·4785	·4419	·3957
·4964	·4649	·4269	·3782
·4855	·4532	·4138	·3628
·4760	·4428	·4022	·3490
·4676	·4337	·3919	·3366
·4602	·4255	·3827	·3253
·4535	·4182	·3743	·3151
·4474	·4116	·3668	·3057
·4420	·4055	·3599	·2971
·4870	·4001	·3536	·2892
·4325	·3950	·3478	·2818
·4283	·3904	·3425	·2749
·4244	·3862	·3376	·2635
·4209	·3823	·3330	·2625
·4176	·3786	·3287	·2569
·4146	·3752	·3248	·2516
·4117	·3720	·3211	·2466
·4090	·3691	·3176	·2419
·3897	·3475	·2920	·2057
·3702	·3255	·2654	·1644
·3506	·3032	·2376	·1131
·3309	·2804	·2085	0

Table XVII

1% POINTS OF z

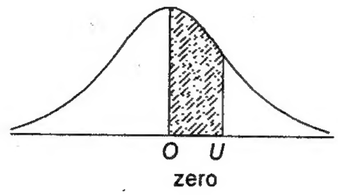
$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	12	24	∞
1	4·1535	4·2585	4·2974	4·3175	4·3297	4·3379	4·3482	4·3585	4·3689	4·3794
2	2·2950	2·2976	2·2984	2·2988	2·2991	2·2992	2·2994	2·2997	2·2999	2·3001
3	1·7649	1·7140	1·6915	1·6786	1·6703	1·6645	1·6569	1·6489	1·6404	1·6314
4	1·5270	1·4452	1·4075	1·3856	1·3711	1·3609	1·3473	1·3327	1·3170	1·3000
5	1·3973	1·2929	1·2449	1·2164	1·1974	1·1838	1·1656	1·1457	1·1239	1·0997
6	1·3103	1·1955	1·1401	1·1068	1·0843	1·0680	1·0460	1·0218	·9948	·9643
7	1·2526	1·1281	1·0672	1·0300	1·0048	·9864	·9614	·9335	·9020	·8658
8	1·2106	1·0787	1·0135	·9734	·9459	·9259	·8983	·8673	·8319	·7904
9	1·1786	1·0411	·9724	·9299	·9006	·8791	·8494	·8157	·7769	·7305
10	1·1535	1·0114	·9399	·8954	·8646	·8419	·8104	·7744	·7324	·6816
11	1·1333	·9874	·9136	·8674	·8354	·8116	·7785	·7405	·6958	·6408
12	1·1166	·9677	·8919	·8443	·8111	·7864	·7520	·7122	·6649	·6061
13	1·1027	·9511	·8737	·8248	·7907	·7652	·7295	·6882	·6386	·5761
14	1·0909	·9370	·8581	·8082	·7732	·7471	·7103	·6675	·6159	·5500
15	1·0807	·9249	·8448	·7939	·7582	·7314	·6937	·6496	·5961	·5269
16	1·0719	·9144	·8331	·7814	·7450	·7177	·6791	·6339	·5786	·5064
17	1·0641	·9051	·8229	·7705	·7335	·7057	·6663	·6199	·5630	·4879
18	1·0572	·8970	·8138	·7607	·8232	·6950	·6549	·6075	·5491	·4712
19	1·0511	·8897	·8057	·7521	·7140	·6854	·6447	·5964	·5366	·4560
20	1·0457	·8831	·7985	·7443	·7058	·6768	·6355	·5864	·5253	·4421
21	1·0408	·8772	·7920	·7372	·6984	·6690	·6272	·5773	·5150	·4294
22	1·0363	·8719	·7860	·7309	·6916	·6620	·6196	·5691	·5056	·4176
23	1·0322	·8670	·7806	·7251	·6855	·6555	·6127	·5615	·4969	·4068
24	1·0285	·8626	·7757	·7197	·6799	·6496	·6064	·5545	·4890	·3967
25	1·0251	·8585	·7712	·7148	·6747	·6442	·6006	·5481	·4816	·3872
26	1·0220	·8548	·7670	·7103	·6699	·6392	·5952	·5422	·4748	·3784
27	1·0191	·8513	·7631	·7062	·6655	·6346	·5902	·5367	·4685	·3701
28	1·0164	·8481	·7595	·7023	·6614	·6303	·5856	·5316	·4626	·3624
29	1·0139	·8451	·7562	·6987	·6576	·6263	·5813	·5269	·4570	·3550
30	1·0116	·8423	·7531	·6954	·6540	·6226	·5773	·5224	·4519	·3481
40	·9949	·8223	·7307	·6712	·6283	·5956	·5481	·4901	·4138	·2952
60	·9784	·8025	·7086	·6472	·6028	·5687	·5189	·4574	·3746	·2352
120	·9622	·7829	·6867	·6234	·5774	·5419	·4897	·4243	·3339	·1612
∞	·9462	·7636	·6651	·5999	·5522	·5152	·4604	·3908	·2913	0

Table X

Areas under the Standard Normal Curve

The area is measured from the mean '0' to any ordinate 'U'.

The table gives the shaded area.



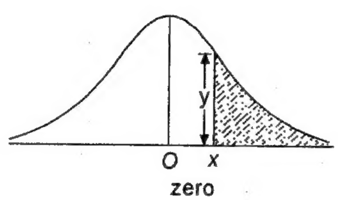
The results are given for values of 'U' at intervals 0·01.

Table XI

Ordinates of the Standard Normal Curve

The table gives ordinates (y) erected at a distance 'x' from the mean *i.e.,*

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

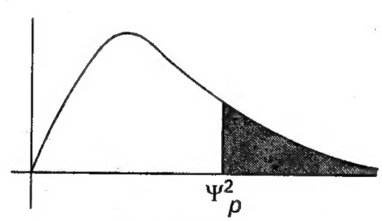


The results are given for values of 'x' at intervals 0·01.

Table XII

Significance Points of ψ^2

The table gives the values of ψ_p^2 for different p's and degrees of freedom 'n'.



Shaded area = p

For large values of n, the expression $\sqrt{2\psi^2} - \sqrt{2n-1}$ may be used as a normal variate with unit variance, remembering that the probability of ψ^2 corresponds with that of a single tail of the normal.

Significant values t_0 of t foı

P_F :

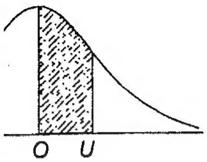
P_F is the shaded area.

v_1 is the number of degrees
the smaller.

v_1 is the number of degrees
the smaller.

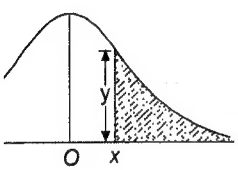
Curve

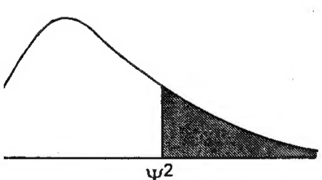
: 'U'.



Curve

om the mean *i.e.*,





Shaded area = p

nay be used as a normal variate
orresponds with that of a single

Table XIII

Values of 'mod. t'

Significant values t_0 of t for given probabilities P_F and *d.f.v.* where

$$P_F = P(|t| > t_0)$$

P_F is the shaded area.

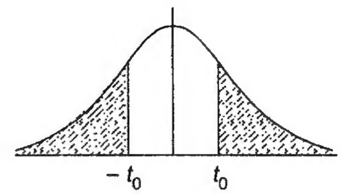


Table XIV

Variance Ratio
5% points of F

v_1 is the number of degrees of freedom for the greater estimate of variance and v_2 for the smaller.

Table XV

Variance Ratio
1% points of F

v_1 is the number of degrees of freedom for the greater estimate of variance and v_2 for the smaller.

❑❑

Index